PFC AND SEMANTIC STRATEGIES

THE ROLE OF THE PFC IN SEMANTIC MEMORY TASKS

By

CHRIS GILBERT, B.A.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Doctor of Philosophy (2008) McMaster University

(Psychology) Hamilton, Ontario

TITLE: The Role of the PFC in Semantic Memory Tasks

AUTHOR: Chris Gilbert, B.A. (McMaster University)

SUPERVISOR: Professor Suzanna Becker

NUMBER OF PAGES: 231

Abstract

The PFC plays an important role in memory tasks in organizing free recall. However, very little is known about the exact mechanisms underlying PFC function. Many researchers, like Morris Moscovitch (1994) believe the PFC supplies cues to other memory areas but details concerning this hypothetical function are vague. Anderson (2003), in contrast, believes that the PFC directly suppresses semantic memory traces. These potential functions of the PFC were explored in the following work. A model of non-strategic memory was built using a TCM framework, and a number of different implementations were evaluated. The model was then applied to Anderson's RIF work, to determine whether an item inhibition account of memory was necessary to explain RIF results. Finally, the model was applied to semantic memory strategies in free recall results to guide empirical research. It was found that no direct inhibition was necessary to explain RIF, and that, in a timed and categorized free recall task, the PFC best performs a semantic strategy by generating category labels at recall. Implications of this work were then discussed.

Acknowledgements

I would like to thank the McMaster Psychology Department for the opportunity to study there, and their financial support. I would specifically like to thank Lee Brooks and Bruce Milliken, who served on my thesis committee, and Sue Becker my thesis supervisor. Finally, both Marc Howard and Ken Norman should be acknowledged for their helpful advice via e-mail.

Table of Contents

The Role of the PFC in Semantic Memory Tasks – Chapter 1

Memory and Organization

Free recall is a demanding test of memory. In a typical free recall task, subjects are presented with a number of memory items during a study session, and, at a later time, they are asked to recall as many of these items as possible. Subjects choose how best to go about this recall process, since they are not provided with any overt cue to use to recall memory items. Even when subjects recall the same number of items on a free recall test, the cues they use and the way they attempt to recall memory items may be completely different. Free recall therefore can be used not only to test the accuracy of recall, but also to examine the methods a subject uses to perform free recall.

Subjects, while performing a free recall task, nearly always organize their recall in some manner (Tulving, 1962). Even when memory items are chosen to minimize pre-existing associations, and when presentation orders are switched up during study sessions, subjects eventually tend to recall study items in a stereotypical order. Tulving created the term "subjective organization" to refer to this kind of consistent ordering of recalls. Subjective organization is often measured by pair frequency analysis (Sternburg & Tulving, 1977). In this method of analysis, memory items that are recalled consecutively are noted. The number of these pairs that reoccur in the next recall session are tabulated, with the total number of these reoccurring pairs representing the amount of subjective organization in the subject's recall. No prior knowledge of a subject's organizational strategy is required to calculate subjective organization. Any kind of organization that leads to a consistent ordering of memory items in recall leads to a high subjective organization score.

Specific methods of organization can also be examined within a free recall task. One kind of organizational strategy in free recall involves the use of semantics. If a free recall task involves memory items that share a semantic relationship, this relationship can be used to organize the items during recall (Bousfield, 1953). Semantic strategies are frequently tested using lists of words (e.g. Allen, Puff, & Weist, 1968). For example, a list of words might contain words belonging to the categories TOOLS, PLANTS, and CLOTHING. During study trials, words might be presented in a random order, with words from each of the categories intermixed with one another. During recall, subjects who used a purely semantic strategy might recall all the words from one category first (e.g. TOOLS) and then words from another category (PLANTS), and then another (CLOTHING).

This kind of organization is often measured using semantic clustering. Semantic clustering involves calculating the extent that words belonging to the same category are recalled consecutively. Crude measures of semantic clustering simply count the number of times words from the same category appear consecutively during recall, whereas more modern measures control for the total number of words recalled as well (Stricker, Brown, Wixted, Baldo, & Delis, 2002). Implicit in these kinds of clustering measures is the assumption that the semantic categories or associations the experimenters assign to list words prior to test and use in the clustering measures at recall are the same categories or inter-item associations that are used by subjects. For example, a subject who recalls some CLOTHING words like sweater and vest consecutively but not others like pants or dresses may be implementing an organizational strategy using the CLOTHING category less than ideally, or may be using an organization strategy using a SHIRT category.

Clustering results should thus only be interpreted in view of the categories or associations used by the subjects, and not in terms of the intended semantic organization experimenters expect.

PFC and Strategies

The pre-frontal cortex (PFC) has been implicated in the kinds of organization in free recall described above. Individuals who suffer from PFC lesions demonstrate significantly reduced levels of recall (Alexander, Stuss, & Fansabein, 2003; Dimitrov et al., 1999; Gershberg & Shimamura, 1995; Hildebrandt et al., 1998; Janowsky et al. 1989; Stuss et al., 1994). For example, in Stuss et al., PFC lesioned subjects performed significantly worse than normal controls on tests of free recall. Normally during free recall, a subject's recall order becomes more and more stereotyped (i.e. recall demonstrates the kind of Tulving-like subjective organization described above) In Stuss et al., subject's recall was shown to be deficient in subjective organization in comparison to controls. The conclusion was that lesions to the PFC impair a person's ability to organize their recall and perform memory strategies.

There are two important exceptions to this finding of deficient memory performance. When a memory task involves information that is already well organized, then PFC individuals demonstrate normal recall. For example, if subjects are presented with information in the form of a detailed story, then when they later try to recall this information, their recall is similar to normals (Janowsky et al., 1989). Also, when memory testing is done in a way that makes retrieval more automatic and less effortful (i.e. when implicit memory is tested), then recall scores are also similar to those of normals. Shimamura, Gershberg, Jurica et al. (1992) tested the memory of PFC

individuals using a word stem completion task, for example. They found no memory deficits whatsoever in comparison to controls.

Besides general organizational deficits, a number of studies have examined the performance of PFC subjects when using semantic strategies. When list materials contain some form of semantic organization, PFC subjects are unable to make use of this semantic information to organize their free recall (Baldo, Delis, Kramer, & Shimamura, 2002; Hildebrandt et al., 1998; Kopelman & Stanhope, 1998). PFC subjects demonstrate significantly less semantic clustering than normals. (But see Alexander, Stuss, & Fansabedian, 2003, for an opposing view that semantic strategies are not disrupted in PFC individuals. This issue is discussed in Chapter 4). PFC subjects do not spontaneously use semantic information to organize their learning and to act as cues during recall. If subjects are given detailed instructions, however, on how to use categories as recall cues and how to try and remember items, then their memory performance resembles that of normals (Hirst & Volpe, 1988). Baldo & Shimamura (1998) examined semantics with a verbal fluency task that is analogous to a semantic strategy. Subjects were given categories to use as cues, and the first letters of to-be-retrieved items (for example FRUIT-a____). PFC subjects performed worse than control subjects. Interestingly, a key factor in PFC subject's worse performance was the large number of repetition errors and perseverations. Dominant responses tended to overcome weaker, contextually appropriate responses.

PFC and Other Facets of Memory

PFC damage can also affect free recall in ways other than strategy use. Some individuals with PFC damage demonstrate overfamiliarity in their recall (Schacter et al.,

1996; Rapesak et al. 1999). These people experience novel events for the first time, but feel like they have experienced them before. Similarly, they can observe a novel object, and experience it as familiar. This deficit may be due to the PFC incorrectly judging memories as familiar. Moscovitch and Melo (1997) extended this finding beyond autobiographical memory. They were able to get individuals with PFC deficits to confabulate historical events and factual information as well as autobiographical events, thus supporting the idea that the cause of this phenomenon was not domain specific, but rather due to some general memory dysfunction.

Patients with frontal damage also can demonstrate the opposite pattern of behaviour – underfamiliarity with actual memories. Levine et al. (1998) found that autobiographical memories can be lost when the PFC (specifically, the inferior, right frontal cortex) is damaged. The way this loss manifests itself is particularly interesting. An individual can have factual knowledge of events that happened to him or her, but may be unable to re-experience these events in any way. In a sense, individuals with this disorder have a limited egocentric view of their own past. This underfamiliarity may be due to a post-retrieval judging process as well. Subjects are able to retrieve memories, but may be unable to use these memories to cue further representations which embody the sensory experiences surrounding the memory.

It has proven difficult to localize the part of the PFC responsible for these deficits, as well as to pin down the exact cause of the deficits. Moscovitch (2002) summarizes the above findings by positing both post-retrieval monitoring difficulties (where individuals have difficulty evaluating generated candidate memories traces) and possible pre-retrieval components as well. Pre-retrieval problems involve the PFC in the memory role

described above: subjects need to select the correct cues to recall memories relevant to the recall situation, and the PFC part of their brain is responsible for this cue selection. It is unclear, then, to what extent over-familiarity in recall is due to the loss of a function distinct from the PFC function implicated in strategic memory. Is over-familiarity due to lesions of a functionally distinct area of the PFC whose sole purpose is to judge the accuracy of candidate memories? Most researchers, as stated below, divide the PFC into separate functional areas. However, while at least one model of the PFC contains a specific mechanism for judging memory suitability (Shallice & Burgess, 1996), many researchers partition PFC function in a way that does not separate this function (Adele Diamond, 2002; Alan Baddeley, 1996 etc.). It is unclear at this time whether some part of the PFC performs a separate post-retrieval judging function, and how the PFC might perform such a function.

The PFC has also been linked to the inhibition of specific memories. Anderson (2003) hypothesizes an "executive control" process which inhibits the amount of activation a specific memory representation undergoes in certain situations. Anderson does not explicitly tie in his executive control process to any specific brain area, but since executive processes are linked so closely with the PFC, his views need to be taken into account when constructing a theory of strategy use and memory. Anderson's views are best understood in comparison to general theories of PFC function. First, an overview of PFC function will be presented, and a general theory of PFC will be extended to explain memory strategies in free recall. Then, Anderson's view of executive function will be presented, and compared to the general PFC theories.

General PFC Theories

There are a significant number of theories of PFC function, some of which focus on how the PFC functions in specific domains such as working memory or problem solving, and some which are couched in more general terms. The literature is complicated by the fact that many of the models differ not only in how they work, but also in what function(s) they perform! Some of the most influential researchers in the field of PFC research are Michael Petrides, Arthur Shimamura, Tim Shallice, Alan Baddeley, and Morris Moscovitch. Below, some of the major lines of research on PFC function are reviewed, although this review will be far from comprehensive, and the research cited here is just a sample of rich, interesting research programs.

Shimamura (2000) introduced dynamic filtering theory. In dynamic filtering theory, the PFC controls the processing of information by performing functions of selecting bits of information, maintaining (keeping information active), updating this active information in some way, and rerouting active processes. Shimamura's research focuses on PFC involvement on memory: semantic knowledge retrieval, episodic recall, and source memory (see Shimamura, 2002 for an overview).

Michael Petrides has produced a significant body of research investigating the functional neuropathways of the PFC. He has run a large number of studies with subjects who had lesions in a particular part of their PFC, and his work attempts to link a particular deficit with a particular part of the PFC (e.g., Petrides, 2000a). He has written numerous book chapters on this topic (Petrides, 2000b; Petrides, 2000c; Petrides, 2002). Some of his earlier work was done on working memory in non-human primates (Petrides, 1994). Goldman-Rakic (1987; 2002) did extensive work on non-human primates as well,

focusing on the working memory capabilities of these animals. Goldman-Rakic also did work on schizophrenia and the frontal lobes (e.g. Goldman-Rakic, 1991).

Baddeley's work, on the other hand, has not been neuropsychological at all, and has been based on the central executive aspect of his overarching memory theory (Baddeley, 1996). Baddeley has focused on tasks, with task switching and dual-task performance studies being key experimental paradigms for him (e.g. Baddeley et al., 2001; Baddeley et al., 1997). He has also focused on Alzheimer's patients as his experimental group, which makes it problematic to generalize his results, since Alzheimer's patients have impairments in many brain areas.

Tim Shallice has embarked on a highly ambitious research program that attempts to model PFC performance across a large number of tasks. He has created the Mark II Supervisory System model (Shallice and Burgess, 1996) as a comprehensive model of PFC performance. The model was complex; boxes, arrows, and functional components abound. In brief, the model dealt with schemas: action plans which compete with one another in particular contexts, and which organize actions once chosen. Shallice's research was also interesting as it includes early and very welcome attempts to compare different models. For example, he contrasted his model with Fox and Das (2000) and found that although superficially the models appear very different, functionally they are very similar (Shallice, 2002). The Fox and Das model is an artificial intelligence model, and the comparison was made to show how two models, developed independently in different domains, hit upon similar solutions when faced with similar problems. This kind of work will be very necessary in the future for theories in the PFC domain, in order to

reduce the large amount of different PFC related terms and functions in the literature into clear, operationally defined, functions.

Jonathan Cohen has developed a computational theory of PFC function designed to address this very problem (Braver, Cohen, & Barch, 2002; Miller & Cohen, 2001). In Cohen's theory, the PFC affects other brain areas through only a single mechanism. It exerts a bias signal to other parts of the brain by reinforcing weaker, contextually appropriate processing or responses to overcome dominant responses or processing. The PFC thus maintains patterns of activity representing goals and means to achieve them, and these patterns result in bias signals to other brain areas.

Morris Moscovitch's research is most relevant to free recall. He has developed a model of memory based on neuropsychological research. It includes an MTL component, a functional area that automatically stores information in conscious awareness, and automatically retrieves information when provided a cue, and a PFC component, which supplies cues to the MTL and supports strategic processes (Moscovitch, 1994). He has also added sensory areas in the posterior cortex as part of his model, since he showed that implicit tests of memory do not depend on the MTL, but rather involve perceptually based associations in the posterior cortex (Goshen-Gottstein & Moscovitch, 1995).

The research programs presented here do not overlap in content to a large degree. Researchers focus on different functions that the PFC hypothetically performs (working memory, planning, inhibition), different subjects (normals, PFC lesioned, schizophrenics, Alzheimers), different tasks (free recall, Wisconsin Card Sorting Task, go-no go, dual task performance, etc.), and different domains (neurophysiology, clinical neuroscience, neuropsychology and cognitive psychology). Rarely do theories compete with one

another in explaining empirical results since research is focused on different areas. PFC research is made more difficult by the fact that the PFC rarely is implicated as the "main" functional area in the performance of a particular task. Because the PFC works in conjunction with other brain areas, researchers not only have come up with different explanations of how the PFC works, but also what it actually does! For example, Shimamura's theory of frontal lobe functions involved selecting, maintaining, updating, and rerouting (Shimamura, 2000); Adele Diamond's developmental research (2002) identified maintenance, inhibition, and manipulation of information, of as well as strategy use, as PFC functions, while Cohen hypothesized that the PFC performs only one function.

Theory of Semantic Strategy Use and PFC

With so many different theories of PFC function, which theory should be used as a framework for explaining semantic strategy use? The obvious choice is Moscovitch's theory of PFC, since it was created to explain exactly these kinds of free recall results. This idea is indeed taken as the framework for the subsequent investigations in this thesis. The PFC is hypothesized to have a cue-providing function to an MTL region which automatically stores and recalls information. However, this theory is modified using some of Johnathan Cohen's ideas (for more details, see chapter 4).

This proposed model of PFC use in semantic strategies differs from Moscovitch's theory in two important ways. In Moscovitch's theory, the MTL module stores memories randomly; intervention of the PFC is required to effectively retrieve memories in particular contexts. In the model proposed here, memories are stored episodically, and "contextual" cues are used by default to retrieve these memories. There are potential

difficulties that arise for any memory system that depends on using contextual cues, though. Specific contextual cues can be difficult to generate if enough time has passed between encoding of the memory and retrieval. A number of memories may be associated with a particular context; also, a number of contexts may be associated with a given event. Memories may be much easier to recall with different cues; in these situations, the PFC is responsible for generating and providing these cues during retrieval.

Secondly, Moscovitch's theory divided the PFC into distinct functional regions. According to Moscovitch, a certain part of the PFC performs the function of providing memory cues, another part monitors recall, and so on (Moscovitch & Winocur, 2002). No such claims are made here. The default position is that the PFC performs one function, that of overriding the dominant "response" of a particular brain area. If a particular part of the PFC is connected to memory areas, then this PFC function results in providing cues. If a particular part of the PFC is connected to brain areas responsible for responding during free recall, then this function results in response monitoring. There are two important caveats to this view. First, there may be many internal mechanisms within the PFC. For example, there may be working memory areas, a central executive, as well as other mechanisms. The theory of PFC function and memory proposed here deals only with the PFC's effect on other brain areas. Secondly, the theory is not dogmatically opposed to multiple PFC functions. The single function view is taken because it is believed to be more useful for a research program investigating the PFC with computer modeling. It is easier to model a single function PFC, and it potentially yields more useful results. Even if a single PFC function can not successfully simulate human performance, a model with such a function will demonstrate why multiple functions are

necessary. The theory proposed here can thus be seen as a combination of Moscovitch's theory of a cue supplying PFC and a "stupid" MTL area with Cohen's general theory of PFC function.

Anderson's Inhibition Theory and the PFC

Anderson's theory of executive function is similar to the theory outlined above. In Anderson's view a stimulus (memory cue) activates a number of associated responses (memory items). When a response is activated above a threshold level, it is "emitted". Since most stimuli activate more than one response, the response that reaches threshold the fastest is the one that is emitted. Stronger activations tend to reach threshold more quickly than weaker ones (although this point is a bit unclear and how activation strength corresponds to response times is not explicitly explained in Anderson, 2003). In order for a weak response to be given for a particular stimulus, stronger responses must first be directly inhibited by executive processes.

Although Anderson's theory is similar to the one presented here, it has one key difference. Both theories posit executive control of interference in a memory process to allow weaker, contextually appropriate responses to "win out" over stronger, prepotent responses. In Anderson's view, this process involves inhibiting the prepotent response, while our theory is explained in terms of increased activation for weaker responses. Functionally, these two explanations may not seem very different. However, Anderson's theory has inhibition that is both specific and long-lasting, and this is significantly different from our theory, as well as any of the theories mentioned above. In our theory, specific brain areas are influenced by the PFC, not necessarily specific representations,

and this influence does not carry over to subsequent tasks. Anderson's theory involves long lasting, item specific inhibition.

Modeling and the PFC

Computational modeling is a good way to explore theories of PFC function in regards to memory. This kind of research involves simulating human performance by using mathematical equations and neuron-like units in lieu of a human brain. Computer modeling can never replace empirical testing, but it can accomplish some things that empirical research can not. Firstly, models can be used in situations where it is impossible to find human subjects. For example, in Chapter 2, a model of human memory is presented in the absence of strategy use. In the real world, subjects do not exist who have no strategy use but otherwise intact brain function, and subjects with PFC lesions may have impairments in areas other than just strategy use. A computer model represents a useful way to explore non-strategic memory, since it easily allows memory simulations where no strategic elements are present. Computer simulations can also act as an existence proof. For example, in Chapter 3, retrieval induced forgetting (RIF) studies are reviewed. A main finding in the RIF literature is that item-specific inhibition is necessary to explain RIF effects. Here, modeling can demonstrate that RIF effects can be explained with alternative theories. Also, the process of modeling can highlight important areas for further empirical study. For example, in examining semantic strategy use in Chapter 4, many of the details of PFC-involvement in semantic strategy use needed to be worked out, and this modeling process lead to new empirical work with interesting findings.

There are a number of different models of PFC function, but the number of models that specifically address the PFC's role in long term memory is very small.

General memory models such as SAM (Raaijamkers & Shiffrin, 1981) and the TODAM models (Murdock, 1993; Murdock 1997) address a large number of different memory findings at once. These models simulate many measures outside of free recall paradigms, and have not been applied specifically to semantic strategy use. Similarly, although there have been attempts to build computational models of PFC function that generalize across tasks, none of these models specifically address semantic strategies. (For a review, see O'Reilly, 2006. In general, computational models of the PFC involve interesting learning rules which explain what motivates the PFC to influence other brain areas, and functional attributes like bistable states and gating mechanisms which explain how the PFC performs its function(s)).

There is at least one model that has specifically examined PFC function in semantic strategy use, however. Becker & Lim (2003) described a model with separate MTL and PFC components. The PFC module in this model used a semantic organizational mnemonic without any explicit training; the model "learned" to use semantic organization in the course of simulated free recall tests. However, human subjects almost certainly make use of existing semantic knowledge when using a semantic strategy and do not learn to perform the strategy itself when performing a free recall task. Also, the semantic cues used by the model were implicit, whereas subjects may be able to use semantic category labels or semantic associations between items to cue recall. So although this model was an interesting explanation of semantic strategy learning, further investigation on the implementation of semantic strategies is required.

Development of the model - MTL

In the majority of real world situations, free recall is a product of the entire brain, including both PFC and MTL regions. In Moscovitch's theory of free recall, the MTL and PFC areas are modular, and each area performs a specific, separate function. In order to model the functions of these brain areas accurately, results are needed that show the performance of one area in the absence of the other. Since the PFC works by supplying cues to the MTL region; it is not possible to measure these cues directly in human subjects. So the best way of building a model of free recall is to first build a model of recall in the absence of PFC function.

Howard & Kahana (1999) examined free recall that controlled for strategic (PFC) influences. In this study, subjects performed a semantic orienting task during the presentation of list words. This task presumably limited strategic processing of list items and rehearsal; subjects had time only to perform the orienting task between word presentations and would not have time to use memory strategies. The study therefore provides results that can be used to develop the MTL part of a hypothetical free recall model.

The results of the study are interesting in their own right. In the study, a number of different types of recency were examined. Recency effects often occur during recall: memory items that are presented at the end of a study session are recalled better than other memory items. If a distracting task is performed between study and test, recency effects disappear, or are attenuated. However, if an identical distracting task is performed between presentations of each study item, recency effects reappear. This pattern of results was found in the study. Also, words that are presented close to one another during study have a tendency to be recalled consecutively. This property of free recall was also

examined in the study, and was labeled lag-recency (for more on these kinds of recency effects, see Chapter 2). A model of non-strategic free recall should be able to simulate these effects, as well as other important free recall results like repetition errors, number of words recalled, and intrusion errors.

Howard & Kahana (2001) built the Temporal Context Model (TCM) to develop an explanation of these different kinds of recency. TCM involves a mechanism of recall called the temporal context. In a simulated free recall task using TCM, words were associated with a pattern of activation in a context layer during study, and were cued with such a pattern during recall. What makes this temporal context layer interesting is how this pattern of activation changes. Whenever a word was presented to the model, an associated context was retrieved, and was used to change the activity level in the context layer. So at a given point in the presentation of a list of words, subsequently presented words were associated with a context that was related to previously presented words. This property of the context layer allowed the model to simulate both recency and lag-recency.

Although TCM provided a novel mechanism of free recall, and provided a comprehensive explanation of recency effects, it was not a fully fledged free recall model. It only simulated the first two recall attempts during a free recall session. It had no way of stopping recall. Building a model of non-strategic memory based on TCM has two advantages, then. First, TCM provides the only complete explanation of recency effects, and it was based on data that controls for strategic influences. It thus makes an ideal base for the non-strategic part of the model. Secondly, in expanding TCM to a full model of free recall, an original contribution can be made. Chapter 2 is an account of the

results found when a number of attempts were made to expand TCM to a full model of free recall.

Development of the model – PFC and Inhibition

Although many neuropsychological accounts of PFC in free recall tasks focus on a strategic, cue-providing function of the PFC, as mentioned above, Anderson's (2003) theory hypothesized an item-specific inhibition function for the PFC. Anderson's work used the retrieval induced forgetting paradigm (RIF). In these kinds of memory tests, subjects are presented with word pairs, and then later practice retrieving only some of them. During cued recall, it is typically found that not only were the practiced word pairs recalled better, but unpracticed word pairs were recalled worse then a control group (where no words were practiced at all). Anderson's explanation for these findings involved a PFC that inhibited specific word representations during practice; this inhibition was theorized to be long-lasting, to carry over to recall, and to be responsible for the RIF effect.

This hypothesized function of the PFC runs counter to the theory of PFC function described above. It is useful, then, to examine RIF effects in a model. If a model can be shown to simulate the main RIF findings without the use of a PFC mechanism, then a useful contribution will be made to the RIF literature, and semantic strategies can be modeled without considering a function of memory representation inhibition. If this kind of inhibition is required, then a model of semantic strategies will need to take Anderson's views into account. Is the PFC involved in inhibiting memories while using semantic cues or not? Under what conditions does the PFC inhibit memories and why?

Only a few changes needed to be made to the model to allow it to simulate RIF findings. First, a semantic layer needed to be added (this would have had to be done regardless to simulate semantic strategies, and so simulating RIF findings is a good intermediate step). Secondly, instead of using context to cue memory items, RIF studies use a cued recall procedure. This cue usage simplified the model, as cues were "fed" to the model during simulated recall. Finally, some means of representing word form was required in order to simulate the presentation of some letters of a word, but not all of it. This resulted in the addition of a word form layer. Chapter 3 describes the results of this model, and its implications for RIF research.

Development of the model – PFC and Semantic Strategies

The PFC has been linked to providing cues during free recall that yield a semantic organization (see above). What are the kinds of cues used in semantic strategies that lead to clustered recall? How and when are these cues generated or learned? Do subjects implement semantic strategies immediately during multi-trial free recall, or does it take time for them to detect an underlying semantic structure? How often will subjects use semantics in a strategy if this property exists in a memory list? Chapter 4 attempts to answer these questions. It involves the implementation of a simple PFC function in the memory model, where the PFC mechanism detects categories, generates them during recall, and uses them as a cue to bias recall to semantics. A number of different potential implementations of semantic strategies are examined in this chapter.

Overview

There are three hypothetical functions the PFC performs in memory tasks. First, it may organize free recall in some manner. A model of memory is developed in Chapter 2,

and applied to semantic strategy studies in Chapter 4 to attempt to clarify many of the details surrounding strategy use in free recall. Secondly, Anderson hypothesizes that the PFC inhibits specific semantic representations during a RIF cued recall task. The memory model of Chapter 2 is applied to RIF studies in Chapter 3 to demonstrate that this hypothetical function is not necessary to explain key RIF findings. Finally, the PFC may be involved in post-retrieval processing during memory tasks. This function is somewhat controversial and unclear; more empirical research should be done to determine when and if executive control is required post-retrieval.

Repetition Error Suppression and Recency in a TCM-like Model – Chapter 2

There is a large number of free recall models, and these models have simulated many important memory findings. For example, the memory model SAM (Raaijmakers & Shiffrin, 1981) has been used to model many facets of free recall, including word frequency effects (Gillund & Shiffrin, 1984), presentation rate and list length (Raaijmakers & Shiffrin, 1980), list strength (Shiffren, Ratcliff & Clark, 1990) and other findings. Other examples of memory models include global matching models like MINERVA (Hinzman, 1988) and TODAM2 (Murdock, 1997), classic generate-recognize models (Anderson & Bower, 1972) and more modern two stage memory models (like Jacoby, Yonelinas, & Jennings, 1997). However, although these models have accounted for many aspects of free recall, no general memory model can explain the full range of recency data (Howard & Kahana, 1999).

The term "recency" used here includes the classic recency finding, long-term recency, and lag-recency. Recency, in its original sense in regards to memory, refers to the finding that words at the end of a to-be-remembered word list are recalled more often than other words on the list (e.g., Postman & Phillips, 1965). If there is a delay between the presentation of a word list and the point at which subjects begin recall (this delay is called the retention interval (RI)), then this delay decreases the magnitude of the recency effect. RIs of over 15s spent while performing a distractor task almost nullify recency (Glanzer & Cunitz, 1966). Long-term recency is evidenced in a memory task where the ability of a distractor task at the end of a memory list presentation to nullify recency is seemingly negated by subjects performing a distractor task between the presentation of successive pairs of list items as well (Bjork & Whitten, 1974). Recency can thus occur

even when there are days between presentation of memory items and recall (Baddeley & Hitch, 1977), so long as there is a significantly large interpresentation interval (IPI) in comparison to the RI.

Lag recency refers to the phenomenon whereby items that are presented close together in a list also tend to be recalled within short lags of one another (Kahana, 1996). For example, suppose a list is presented containing the word sequence CAT, DOG, MOUSE, FROG in order. The word CAT would be a lag of -1 away from DOG, and the word MOUSE would be a lag of +1 away from DOG, while FROG would be +2, etc. Words that have smaller lags between them have a higher probability of being recalled next, given that one of the words has just been recalled. For example, if a subject had just recalled DOG, then there would be a greater chance of MOUSE being recalled next, in comparison to FROG, because MOUSE is a lag of 1 away, while FROG is a lag of 2. Lag recency is also asymmetrical in that word pairs with a positive lag tend to be recalled together with a greater probability than word pairs with an identical negative lag. If DOG had just been recalled from the example list, then MOUSE would have a greater chance of being recalled next, in comparison to CAT, a word with a lag of equal magnitude but opposite sign.

Lag recency findings have been difficult to simulate using a general memory model (Howard & Kahana, 1999), and instead, have been explained in terms of a new memory mechanism included in Howard & Kahana's (2001) Temporal Context Model (TCM). TCM contains three main components – a representation of memory items, a representation of context, and numerical matrices (analogous to weights used in connectionist modeling, and hereafter referred to using this term) representing the

strength of association between memory items and context. During the simulation of the study phase of a free recall task, items are associated with the current pattern of activation representing context by changing weight values, and during the simulation of testing, a contextual pattern (cue) is modified by the weights to generate potential retrieved items. TCM thus has a very simple structure, and involves simple learning mechanisms, similar to those found in many memory models. What distinguishes TCM from other mathematical and connectionist models of memory, and allows it to simulate recency results, are two significant properties.

The first important property of TCM is a kind of contextual drift that allows the model to simulate lag recency results. Initially, a random vector of unit length represents context prior to study. As memory items are "presented" to the model, each memory item is first associated with the current contextual pattern, and then changes this pattern according to pre-learned contextual associations (retrieved context). This gradual contextual drift causes memory items that are presented close to one another at study to have more similar contextual associations than items presented farther apart, because of context changes that occur over time due to context retrievals. During simulation of testing, the contextual pattern at the end of study is used to cue memory items. This pattern will be most similar to items at the end of the memory list, and so recency is simulated. Once an item is selected and given as a response, this item modifies the context pattern according to learned contextual associations (retrieved context again). This new contextual pattern then cues memory items, and since this new contextual pattern (partially) represents the contextual associations of the just-remembered-item,

those items presented closest to the just-remembered-item are most likely to be recalled. Thus, lag-recency is simulated.

Long term recency effects occur in TCM due to a competitive retrieval process, the second important property. When items are cued by the context layer, a number of different memory items have significant activation levels. Some rule or mechanism is required to choose amongst these candidate items. Howard and Kahana (2001) show that if a rule is used that selects items probabilistically according to their relative activation strength, then long term recency can be simulated. A probabilistic selection method selects an item according to how strongly activated it is compared to all the other memory items Thus, it isn't the absolute strength of a memory item that determines whether or not it will be given as a response when it is cued by the context, but its relative strength. If a long delay weakens all items to the same degree, then there will be no effect on recency, because the relative strength of a particular item in comparison to other items will be unchanged. This is how long term recency is modeled in TCM.

A single process explanation of recency, as found in TCM, is not uncontroversial. Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, and Usher (2005) question the use of a single process explanation due to dissociations between short and long-term recency (such as those found in directed output order tasks, amnesiac patients, and negative recency results). They have constructed their own model of recency which includes an activation-based short-term memory buffer, and a weight based, episodic memory modified by a random walk contextual mechanism. The following work does not attempt to adjudicate between single process and two process explanations, but rather, attempts to

examine how a single process explanation (using a TCM-like retrieved context) influences other key free recall results.

TCM is not a complete model of free recall, because it lacks the necessary control mechanisms to simulate multiple recalls. It could instead be described as a test of a hypothetical mechanism important to recency results. TCM only models the first response of free recall, and the probability of a word being given as the second response. It does not model subsequent recall attempts. One reason it is unable to simulate repeated recalls is that the model has no method of stopping recall (and does not need one since only one the 1st recall attempt is simulated). TCM also models only a particular kind of free recall data: one trial free recall in which a semantic orienting task is used (Howard & Kahana, 1999). The semantic orienting task is a control used to limit the effect of rehearsal and strategic processing, since both of these things can interfere with recency and lag-recency results. Finally, even within this limited domain of free recall tasks, not all memory results are simulated, only those specific to recency. For example, during free recall of word lists, a small but significant amount of the time repetition errors occur. TCM has no method of preventing repetitions at all, and the model might give as output the same memory item many times in a row. So the wide range of free recall results that models like SAM and TODAM can account for cannot be accounted for by TCM. An ideal memory model would have the breadth of classic free recall models along with the ability to account for recency effects on par with TCM.

To build such a model, one could either try to extend a current general memory model to simulate recency results, or build a model based upon TCM principles, extended to incorporate the broader set of features required to model full free recall. It is not clear

how to extend current mathematical models to handle recency and lag-recency, and it has proven difficult using modeling methods outside of TCM (Howard, personal communication). To make TCM into a model of free recall at least two things must be done. A mechanism of preventing repetition, and a mechanism of stopping free recall must both be built in.

There are two main classes of repetition prevention mechanisms in the modeling literature. Response suppression (Burgess & Hitch, 1999; Farrell & Lewandowsky, 2004) involves eliminating an item as a candidate for response once it has been given in recall. This can be done by means of a formal rule that ignores already given items, by temporarily changing the weights connecting to the item, which will cause it to be less active, or by actively suppressing activity in units that represent the item. Burgess & Hitch (1999) have hypothesized a form of response inhibition in which the same item is not given as a response in successive or proximate recalls due to neuronal fatigue. Response errors do occur sometimes, however, and so any response prevention mechanism needs to allow these errors to happen a small percentage of the time. Response suppression mechanisms can decay slowly over time, allowing for repetition errors to happen infrequently over a large enough time interval. If a response suppression mechanism is used, an additional mechanism must also be implemented to tell the model when to stop. Human subjects do not try to recall words indefinitely. In some experimental tasks, the experimenter will cut off recall, but human subjects are, of course, capable of stopping recall without outside intervention.

The second class of repetition prevention mechanisms involves a memory model with a two-step procedure such as generate-recognize (e.g. Anderson & Bower, 1972).

The first phase of these kinds of models involves the activation of candidate memory items. This activation is followed by some kind of process that judges the suitability of the generated item as a response. This second stage allows for the prevention of repetitions. Once an item has been given as a response, the model can recognize this fact when the item is generated a second time, and reject it as a response. Repetition errors can occur due to noise in the recognition mechanism, or due to decay in the association between an item and some contextual representation of the fact that it has already been given as response. The model can use a certain number of rejected generations as a cue to stop recall. A generate-recognize mechanism can also act as a mechanism for stopping recall.

TCM is being extended by Sederberg, Howard and Kahana into a model of free recall called TCM-A. This extension of TCM is a more substantive version of the horse race model described below, and is being developed independently of the models described here. The following work is not an extension of TCM per se, but rather an examination of the retrieved context mechanism of TCM in a number of different models.

The purpose of the models was to compare different methods of repetition prevention and recall stopping in a retrieved context framework. It is not clear how these additional mechanisms will interact with the key components of TCM: contextual change by retrieved context and probabilistic response. Also, although probabilistic response selection and contextual change by retrieved context allow TCM to simulate recency effects, they may have other negative effects on other key free recall results. The goal of this paper is therefore to explore the important properties of TCM in a full model of free recall. Various methods of stopping recall and preventing repetition will be implemented

in a TCM-like model, and these models will be evaluated to determine which most accurately simulates human data, and how the choice of these mechanisms influences other free recall results. Also, a wide range of free recall results will be simulated to demonstrate how a temporal context influences other key free recall findings. Two main questions will be addressed: 1. What computational principles are required to build a TCM-like model of free recall? 2. Do these principles interfere with other important free recall findings?

Method

Current Model vs. TCM.

The model described below differs from TCM in a significant number of ways. It is easier to describe how the two models are similar than to outline these differences. The purpose of the model was not to implement TCM, but rather to implement the ideas contained in TCM that allow it to simulate recency effects, in a model of full free recall, and in as simple and biologically plausible a manner as possible. The "important" parts of TCM are its retrieved context and a competitive retrieval process because these mechanisms allow it to simulate the full range of recency effects (its main initial contribution to memory research.) A context that changes over time during study allows a model to simulate lag recency, and if the context change is due to word retrievals like in TCM, then these lag recency results are asymmetrical. A competitive retrieval mechanism allows for long term recency, since in these mechanisms it is the relative strength of items rather than the absolute strength of items that is important. When the model is described as TCM-like, it is a way of saying that it has these two properties. Any other similarities between the below model and TCM are unplanned.

Overview of the Model

Before describing the model in detail, an overview will be given. The model contained a context layer and item layer. Each unit in the item layer represented a potential memory item. The context layer started with a random contextual pattern. Items were "presented" to the model by causing a pattern of activation representing an item to appear in the item layer. Weights between the item layer and the context layer were updated by a simple Hebbian learning rule. At the same time, each item had a pre-existing context associated with it, and this context was used to change the current pattern in the context layer so that it was more similar to this retrieved context. A new item was then presented to the model, and it was associated with the new contextual pattern. This process continued until all study items had been presented.

At recall, the current pattern in the context layer was used to cue the item layer, using the weights between the two layers. This resulted in pattern of activity in the item layer. A particular unit (representing a memory item) was selected using a softmax decision rule. In generate-recognize versions of the model, a recognition decision was made on this selected unit. If the unit was not recognized, the context pattern was used again to cue the item layer. If the unit was recognized as a list item, it was said to be given as output. In response-suppression versions of the model, the selected unit was always given as a response; this unit then received a negative bias making it less likely to be given as a response in the future. After an item had been generated, the contextual pattern used to cue items was changed to become more similar to the existing contextual pattern associated with the just recalled item. The model continued to attempt to recall

words using the current contextual pattern as a cue until a stopping threshold was reached. The context layer was later split into two separate parts (see below).

Basic Model

The base model was composed of two layers, shown in Figure 2.1: a memory layer, and a temporal context layer. The memory layer stored representations of the to-be-remembered word items. Each unit in this layer represented one word; for example, the word DOG would be represented by a particular unit having an activation of 1 and the rest of the units having an activation of 0. This unit would only be active for DOG, thus the model used a localist representation of words. The temporal context layer contained a distributed pattern of activation representing a temporally changing context, similar to TCM (Howard & Kahana, 2001). At any point in time, this layer had units with activity levels between 1 and 0. All of these units taken together (the activity vector) represented the current context of the model. All layers were implemented as rate-coded neural circuits and there was full interconnectivity between layers.

Training

Prior to training, every word unit in the memory layer corresponding to a memory list word was associated with a context vector in the context layer (the weights were set so that a value of 1 in a unit in the memory layer exactly retrieves this context vector). The pre-list context vectors were randomly generated vectors that were mutually orthogonal to one another (i.e. they had a unit length of 1 and they were all orthogonal to one another). These pre-list context associations were meant to represent an average of the temporal contexts of all the times a particular word has been presented to the model. This simulates a human's prior experience with list words prior to testing. Additionally,

non-list word items were also given pre-training contextual associations. The contextual patterns that were trained were mutually orthogonal to one another and the contextual patterns of the list items. For a few of the intrusion items, these contextual patterns were then modified with a very small amount of random noise, causing them to be slightly non-orthogonal. Before training commenced, the context layer was set to a random vector, representing the context right before testing begins.

The model was trained to simulate the study phase of a subject's learning during the spoken presentation of a list of 12 words during a free recall task. Each word 'item' was presented to the model by setting the activity of the word unit in the memory layer that corresponded to the word item to 1, and the activities of all other units in this layer to 0. This active unit was associated with the current context (the current pattern of activation in the context layer) by applying the following Hebbian learning rule to the connection weights between the two layers:

$$\mathbf{W}_t = l(\mathbf{c}\mathbf{i}^T) + \mathbf{W}_{t-1} \qquad (1)$$

where $\mathbf{W}$ is the weight matrix between the context and item layers, l is the learning rate, $\mathbf{c}^T$ is the transpose of the current context vector, t is the current time step and $\mathbf{i}$ is the item layer activation. Note that lowercase letters denote vectors, and bold uppercase letters denote matrices.

Each time this Hebbian learning took place, the context layer updated its context representation. This new context was a function of the old context activation and a retrieved context:

$$\mathbf{c}_t = f\mathbf{r}_i + \sqrt{(1-f^2)}\mathbf{c}_{(t-1)} \qquad (2)$$

where f is a context change parameter and $\mathbf{r}_i$ is the retrieved context for item i. The retrieved context was calculated by multiplying the active word unit's activation with the weights between this unit and the temporal context layer:

$$\mathbf{r}_i = \mathbf{W}\mathbf{i} \tag{3}$$

where i is the pattern of activation of unit i, and **W** is the weight matrix between the item and context layers. Once the new context was instantiated in the temporal context layer, the next item on the list was presented to the model, and training continued until all 12 list items were presented.

The performance of a distracter task during recall was simulated by changing the context in the following manner:

$$\mathbf{c}_t = f_2\mathbf{c}_{random} + \sqrt{(1-f_2^2)}\mathbf{c}_{t-1} \tag{4}$$

where **c** is the current context, t is the current time step, random refers to a random unit length vector, and f_2 refers to a context change parameter. The model did not simulate the performance of the distracting task, but rather, its effect on subsequent recalls. The distracting task had the effect of causing random contextual drift.

Recall

A subject's recall of a list of words in a one trial, delayed free recall task was simulated as follows. The activity level in the temporal context layer at the end of training carried over to the start of simulated recall. In this case, activity level was modified by randomly adding noise to the temporal context, thereby simulating a delay between study and recall:

$$\mathbf{c}_r = \mathbf{c}_s^T \times l\mathbf{v} \tag{5}$$

where $\mathbf{c}_r$ is the context at the beginning of recall, $\mathbf{c}_s$ is the context at the end of study, l is a value between 0 -1 that is closer to 1 the longer the simulated delay, and $\mathbf{v}$ is a random unit vector. This additional noise is not required to simulate the free recall results shown below, and can be omitted.

Simulated recall began by having the activation in the temporal context layer cue the memory layer:

$$\mathbf{i} = \mathbf{W}^T\mathbf{c} \tag{6}$$

This cuing resulted in the word units in this layer being activated by a certain amount and the model "choosing" amongst these activated units using the softmax function (Brindle, 1990). The softmax function selected one of the active units in the memory layer and set its activity level to 1 and at the same time it set the other units activities level to 0. This selection process was done probabilistically, so that units with a greater level of activity were more likely to be chosen than less active units:

$$p_i = \frac{e^{i\mu}}{\sum_{j=1}^{n} e^{j\mu}} \tag{7}$$

where p is the probability of item a being selected, j is an item in the item layer, i is the activation of a particular unit, n is the number of items in the item layer, and μ is the softmax parameter.

Softmax selection simulates the activity of inhibitory lateral connections in memory areas in the brain and their effect on competing activity in these brain areas (Brindle, 1990). The softmax function is thus a divisive form of lateral inhibitory similar to Luce's choice rule. It should be noted that the exact function of inhibitory neurons is

disputed, and that the softmax function represents a simple model of one hypothetical function.

Generate Recognize Model

Two different methods were employed in the model to prevent repetitions and to stop recall, which we will refer to as the generate-recognize (GR) model and the response suppression model. First, the GR model will be explained. The base model of memory was extended into a GR model by the addition of a recognition decision process. This recognition decision was performed on items generated by Equation 6. The generated word's retrieved context was calculated according to Equation 3. This retrieved context was compared to the current context (the pattern of activation in the context layer that was used to cue the item layer.) The generated word was said to have been recalled if

$$z_1 < |\mathbf{Wi} - \mathbf{c}| < z_2 \qquad (8)$$

(where z_1 and z_2 were upper and lower thresholds for rejection, $\mathbf{c}$ was the activation in the context layer, and $\mathbf{aw}_a$ was the retrieved context). This equation allowed the model to reject candidate items if they are too similar (i.e. a repetition of the item just recalled) or too dissimilar (intrusion errors) to the cue used to generate them, so that immediate repetitions were prevented, as well words judged to be to dissimilar to the cue used to generate them.

If a word unit in the memory layer was recalled successfully, the temporal context layer updated using Equation 2. This new context then cued the memory layer, and recall proceeded. Recall ended after a certain number of rejected generations. The model was run a number of times equal to the number of simulated subjects For the full set of modeling parameters, see Appendix F.

The GR model described above does not contain any additional structural components, and merely adds an additional recognition stage to the existing base model. However, simulation results were poor with this model (see below). An additional component was therefore added to the model. This second generate recognize model (GR2) split the context layer into a slowly varying "constant context", and a rapidly varying "temporal context". The temporal context layer acted identically to the context layer described above. The constant context layer had no pre-study associations. During study, a pattern of activation (consisting of a vector made up of random values between 1 and 0) representing the study context was instantiated in this area, and word items were associated with this pattern, using Equation 1. This pattern of activation stayed constant over the simulated presentation of list words. During test, a pattern of activation representing a test context was instantiated in this area. After the successful retrieval of a list word, the retrieved word would be associated with this new pattern. Recognition decisions were made with this layer as well; items that were associated strongly with the current test context were rejected as repetitions. This form of the model is similar to models by Vousden and Brown (2000), where contextual changes occur at varying rates.

Response Suppression Model – Training and Simulation

The response suppression model was identical to the generate-recognize model except during simulation of recall. During recall, the response suppression model had no recognition phase. Any word unit that was activated with the softmax function was given as the model's output. Once a word unit had been activated, its activity level was influenced by a bias unit with a negative value. For subsequent recalls, the activity level of each unit was equal to the cuing from the context layer, as well as this negative bias:

$$i = \mathbf{W}^T \mathbf{c} + b \tag{9}$$

where b is a bias parameter, and $\mathbf{W}_i$ is the weights between the just activated unit i and the context layer. This equation was applied to each unit in the item layer to form an activity vector, and the softmax function was then applied to this vector using Equation 7. The result of this negative bias was to cause the word unit to be unable to be activated on recall attempts after it has been given as a response. The amount of negative bias decayed over time. For a given unit i,

$$b_t = b_{t-1} d \tag{10}$$

where d is the rate at which the bias decays.

This allowed for some repetition errors to take place with increasing probability as the number of subsequently recalled words increased.

The response suppression model continued to try to recall words until the activity level in the strongest item was below an absolute threshold z (before the softmax function was applied). If the most highly active unit in the item layer exceeded this threshold, the model stopped recall.

Horse Race Model – Training and Simulation

In order to simulate temporal dynamics, another response inhibition type model was designed. Like the response suppression model, it was made up of a context layer and a word unit layer, with identical connectivity and structure. The study phase of this model was identical to that of the response suppression model. However, the horse race model had an additional output layer that was used during simulated recall, and each individual recall attempt was broken down into several discrete time steps (see Figure

2.2). This kind of model is similar to accumulator models like RACE (van Maanen & van Rijn, 2006).

During recall, the horse race model has the temporal context layer cue the word unit layer with a modified version of Equation 5:

$$\mathbf{i} = \mathbf{W}^T \varpi \mathbf{c} \tag{11}$$

Where ϖ is a time step parameter.

This activation is then sent to the output layer where it accumulates across time:

$$o_t = o_{t-1} + i_{t-1} + \eta \tag{12}$$

where η is a random noise parameter, o_a is the value of the output unit for item a, and i_a is the value of the item unit a.

The model continuously performed these two equations until a unit's activation in the output layer exceeded a threshold. Once an output unit exceeded the threshold, this unit's output (representing the word unit it is connected to) was given as the response of the model. The activation level of this unit was then biased for subsequent recalls according to Equations 7 and 8, in a similar fashion to the response suppression model. The model continued to recall words in a similar fashion, and ended recall once the pre-softmax activation level in the word units was below a given threshold, as in the response suppression model. A time-based threshold was also used in one version of the model. In this case, if an activation level in the output layer did not exceed threshold for output in a given number of time steps, then the model ceased to attempt recall.

Results

Several simulations were run with the three models described above, with the following goals:

- To replicate the original TCM model's ability to simulate probability of 1st recall and lag-recency results using our simple, retrieved context model, and incorporating the two main properties of TCM: contextual change by retrieved context, and probabilistic, competitive recall
- To arbitrate between different kinds of repetition prevention mechanisms in a TCM-like model by examining temporal dynamics, total number of words recalled, and repetition error performance
- To test TCM's viability as a general model of free recall by extending it to simulate such phenomena as intrusion errors, and multi-trial free recall.

All the response suppression results were based on the response suppression model described above, under that subject heading. Although the horse race model is a kind of response suppression model, in the following sections it will always be referred to as the horse race model.

Lag Recency and Probability of 1st Recall

TCM was developed to accurately simulate lag-recency and has done so very accurately in past work (Howard & Kahana, 2001). It may therefore seem redundant to simulate these results again. However, the above models and TCM differ in several crucial aspects. First of all, the above models simulate a full free recall session, whereas TCM only examines the 1st word recalled, and the subsequent probabilities of potential second words being recalled from a list. Thus, it is unclear if lag-recency effects can be simulated for words recalled after the first. Secondly, the above models contain mechanisms that stop recall and that prevent repetition. These mechanisms might potentially influence lag-recency results in some manner. Finally, although the above

models use contextual change by retrieved context, other aspects of the model are distinct from TCM, and may influence results.

The simulation results for probability of 1st recall are shown in Figures 2.3, 2.4 and 2.5. The human data came from a Howard and Kahana study (1999) of delayed one trial free recall, where (as mentioned above) a semantic orientating task was used to minimize rehearsal and strategic processing. There were three conditions used in this study: an immediate condition, where testing occurred immediately after study (Figure 2.3), a delayed condition, where testing occurred only after an arithmetic distracter task had been performed for at least 10 seconds (Figure 2.4), and a continuous distractor condition, where a distracting task was performed between item presentation during study, as well as between study and test (Figure 2.5).

There was no significant difference between the generate-recognize and response suppression models in simulating probability of 1st recall. There was a greater chance for words presented last during study to be recalled first in the immediate and continuous distractor conditions, and this effect was attenuated in the delayed condition. Thus, both models accurately simulate the performance of humans on this measure. Delayed free recall attenuated the recency effect in the model because the delay task after study causes the context to drift, and thus become less similar to the context associated with items at study. The recency effect returns in the continuous distractor condition because contextual drift (due to a distractor task) occurs between presentations of each successive pair of study items, and because it is the strength of an item relative to competing items that causes it to be recalled. The absolute association strengths of each item to the initial context at test were much lower than in the immediate condition; however, the relative

strengths were similar due to the distracting task performed between the presentations of each item during study.

The lag probabilities are shown in Figures 2.6 and 2.7. Initially, the generate-recognize model did not yield correct lag probabilities. This result was due to the models not stopping their free recall properly. Recall either went on indefinitely, or for a highly variable amount of time. In the models, retrieval of a given item results in a change in the context that is used to cue subsequent items. Any method of stopping recall that is based on this context, or item strength that is derived from this context using Equation 6, runs into problems since this contextual change is not entirely predictable. All subsequent results for the generate recognition model were based on the modified version (GR2) of the model described in the method section. Lag probability results were close to identical for the generate-recognize and response suppression models. There was a greater probability of a word being recalled over smaller lags, and there was an asymmetrical effect for positive lag, in the performance of both the models and human subjects. (Note that lag probabilities were calculated as the probability, given a response of item x, that an item with a lag of y from x would be subsequently recalled, as long as there was a candidate item with that lag. So, for example, if the previous word recalled was the last item on the list, there would be no list words with a lag of +1 from that item, and so this particular situation would not be included in calculating the overall +1 lag probability score.) To understand how the temporal context achieves these lag results, see Howard & Kahana, 2001. The model also predicts a greater lag asymmetry on the first several recall attempts. This effect was due to recency. The last items on the simulated list had an increased chance of being recalled due to a context that was still similar to that of the

context that was used on the first recall trial, where recency effects were found. In other words, initially contextual cues cause items later in the list to be recalled with a higher probability, and this initial contextual pattern slowly becomes "erased" over time due to retrieved contexts. The presence of a repetition prevention mechanism and a recall stopping mechanism did not influence lag probability results directly. When simulating lag-probability results, however, issues with stopping mechanisms used by the response suppression model were discovered. This will be discussed further in the next section.

Stopping Recall

As mentioned above, the models initially performed inadequately at stopping recall. When the generate-recognize model was modified (see method), it was able to stop recall and simulate total number of words recalled. The response suppression model was able to stop recall but had the issue described below.

The generate-recognize model stops recall after a given number of unrecognized generations. Figure 2.11 shows the model's performance in comparison to human subjects. Once again, human data is based on Howard & Kahana's (1999) one trial delayed free recall condition. Two groups of subjects from that study were used, one group that had a 10 second delay (retention interval) between study and recall, and one group that had a 16 second delay. The model performs similarly to human subjects at delayed free recall after a 10 second delay after study. Importantly, if the delay is increased to a 16 second retention interval, the model still performs like human subjects without any parameter changes. Interestingly, the model recalls more words at greater delays because of a decrease in recency. The more a set of items are associated with a given temporal context, the greater chance that these items will be given as a response

and then generated again and rejected. Since the model stops after a certain number of rejections, a greater recency effect is linked to fewer words recalled (unless the contextual change parameter is very high).

Human subjects seem to voluntarily stop the recall of words in free recall. Subjects tend to stop recall after about 10 seconds of unsuccessful free recall (Wixted & Rohrer, 1994). However, if subjects are asked to continue to try to recall more words after they have stopped, they are able to do so, after progressively longer and longer delays. These results fit in perfectly with the generate-recognize model, because generations are the model's measure of time. Since the model always stops after a certain number of incorrect generations, this parallels human subject's tendency to stop after a 10 second time period of no recall. As with human subjects, if the threshold of incorrect generations (time) is raised, the model is able to recall more words.

The response suppression model uses an absolute threshold based on the activation strength of word units in the item layer; this threshold is applied to the total activation prior to the application of the softmax activation function. If the most strongly activated word unit was not above this threshold, then the model stops recall. This stopping method has the disadvantage of being difficult to interpret in terms of brain processes. If the softmax function simulates lateral inhibition between competing memory representations then it is questionable to assume that these representations can be active without this lateral inhibition influencing their activity levels. Having a model stop based on the activity levels before the simulated lateral inhibition is questionable.

The effects of this absolute threshold are shown in 2.12. The model simulates the performance of human subjects at a retention interval of 10 seconds. After a 16 second

retention interval, the model does not recall any words whatsoever. This is because after longer delays, the temporal context changes to a greater degree from its representation at the end of study. Thus, when this temporal context cues words, the absolute activation levels are lower then they normally would be, and in this case, they were actually below the threshold for stopping recall. To address this, the threshold could be raised so that recall didn't stop, or the degree of random fluctuation in the temporal context between study and recall could be lessened, to allow the model to recall words at this greater retention interval. The model still performed worse over longer retention intervals, whereas human subjects perform better after a 16 second retention interval in comparison to a 10 second retention interval. So regardless of the values used, the response suppression mechanism could not simulate the total number of words recalled as a function of different retention intervals.

Repetition Errors

Response suppression models were able to simulate repetition error data from human subjects (see Figure 2.9). The human data was taken from an analysis of Lamming (2005) of a one-trial free recall study (Murdock & Okakda, 1979). An erroneous repetition of a studied word tended to take place four or five recalls after the word was given as a response the first time in both human subjects and simulated subjects. Also, the overall number of repetitions across lags for all 72 subjects was modeled accurately. This result was not surprising, since a specific kind of decay was implemented in response suppression specifically in order to model this result.

Generate recognize models simulated these results without requiring additional mechanisms (Figure 2.8). The model did not repeat items because they were associated

with a pattern representing testing in the constant context layer. Immediate repetitions were prevented due to the properties of the recognition stage: items that were too similar to the context that cued them were rejected. However, if a recognition decision was based on both the constant and temporal context layers, and the threshold for rejection was not set too high, then a small percentage of the time an item will be incorrectly given as a response because its temporal context component was similar enough to the contextual cue used to warrant an acceptance of the item as a response. This method of preventing repetitions had the advantage of not requiring specific mechanisms to fit the data, and it fell out of the general architecture of the model. This method prevented repetitions on the 1st recall attempt after an item has been recalled, and repetitions peaked at the 4th or 5th recall attempt after a given item has been recalled; this closely mirrors the performance of human subjects on one trial free recall of lists 20 words, who typically do not recall more than 5 or 6 words (e.g. Gilbert & Becker, in preparation).

Temporal Dynamics

The generate-recognize model simulated temporal dynamics accurately (see Figure 2.10). As subjects recalled words on a free recall task, they tended to take a longer amount of time for each subsequent word. Their cumulative recall scores plotted across time yield an exponential function (Wixted & Rohrer, 1994). The generate-recognize model simulated this result by assuming each word takes the same amount of time to generate. As more words were recalled, it became increasingly probable that an already-recalled word will be generated again. This repetition was usually correctly rejected by the context layer. Since these rejected generations took time, on average it took more

time to generate a correct, not-already-recalled word later in recall then it did for earlier recalls. This lead to the temporal dynamics shown in Figure 2.10.

For response-suppression models, there was no explicit measure of time built into the model, and there was no a priori reason to believe that later recalls would take longer than earlier ones. The response suppression model's cumulative number of recalls across time was therefore a straight line, which is very different from the performance of human subjects.

In order to attempt to simulate the temporal dynamics of free recall using a response suppression method, the horse race model was introduced. Here, the time it takes an item to be recalled depended on the amount an item was activated in the memory layer, the amount of noise used, and the threshold used. The horse race model was either able to successfully model temporal dynamics but not recency data, or recency data but not temporal dynamics. If there was no noise built into the model, then the item that was most active after being cued by the context layer always won. If the noise level was very high in comparison to the level of activity in the memory layer, then the chance of each item being recalled was nearly identical to the chance of every other item being recalled. In order to simulate the previous lag-recency and probability of 1st recall results, the amount of noise used in output calculations had to be very precise. Unfortunately, no noise values in this range lead to correct temporal dynamics.

It might be possible to simulate both temporal dynamics and lag-recency using the horse-race model if different parameter values were used during training (such as initial weight values, learning rate, and contextual change parameter). However, in our simulations, sampling a wide range of parameters, no combination was found that could

simulate both temporal dynamics and recency results. The parameter space was searched thoroughly, using a large sample of parameter values, calculating the best fit with the data, and then using the best parameter values as a starting point for a subsequent, similar parameter search. We speculate that this negative finding was due to two properties of the horse race model. First, the model was extremely fragile regardless of the parameters used during training. There was always only a very small range of potential noise values that can be used to get lag-recency and recency results; in fact, no set of parameters were found that lead to accurate simulation. This property of the model made it unlikely to be the functional basis for temporal dynamics in memory. Secondly, regardless of the amount of noise level used, the model had to output words at a slower and slower rate, and have this output time follow a hyperbolic function. Since noise was at a constant value across all recall trials, the average amount of time to recall words at a given trial was dependant upon the item strength. This meant that the context layer would have to act as a progressively less effective cue for the item layer, so that items, on average, would have a lower strength and thus take a longer time to recall. The model simply did not perform this way unless the weights between the context and item layer decayed using a hyperbolic decay function. There is no reason to reason to believe this is the case.

It might also be possible to model temporal dynamics in a response suppression model if lateral inhibition and item selection were explicitly modeled in a detailed biophysical model, rather than modeled in a more abstract form using the softmax function. If lateral inhibition was modeled in a manner where, for example, the time it took an item to be recalled was based on the absolute strength of the items, but the chance an item was recalled was based on its relative strength in comparison to other items, then

both temporal dynamics and recency might be modeled using response suppression. Problems arise in the current version of the model using the "horse race" at the response stage because both recency and temporal dynamics depend on the same measure.

Multi-trial Recall

Multi-trial free recall almost always involves some sort of organizing factor or strategy, and so in general, multi-trial free recall results are beyond the scope of this paper. However, retrieved context can be shown to give a very important property of multi-trial free recall. If subjects study a list of words on more than one occasion, and are tested after each study session (each study-test procedure is a trial), then the number of words a subject recalls increases after each study session. This model can partly simulate this finding. For example, after one study trial, the generate-recognize model recalled 4 words successfully. After another study session, the model then recalled 7 words successfully. More words were recalled after each study session for two reasons. First, there was less of a recency effect on the second recall trial, and secondly, lag-probabilities were higher for greater lags, and lesser for shorter lags on the second recall trial. Since correct rejections cause the generate-recognize model to stop, if all the list words have recall probabilities that are closer together, there was less chance of an item being generated more than once, and so less chance of the model producing the threshold number of correct rejections. However, this increase in words recalled depended crucially on the association strength of extra-list items to context, and other parameter settings. Also, learning did not continue past the second recall trial. In conclusion, recency and lag-recency results decreased significantly in multi-trial recall, a universal

consequence of a temporal context, and in some instances this decrease led to more words recalled.

Intrusion Errors

As shown in Figure 2.13, the generate-recognize model simulated intrusion errors. Occasionally in the course of free recall, subjects will give as a response a word unrelated to any of the list items. In the model, a large number of these "words" were constructed,. One of these intrusion words was incorrectly recalled very rarely, similar to human subjects. The rate at which intrusions were given depended upon the number of potential intrusion words, and the overlap between the current context and pre-existing contextual associations of the intrusion words.

Discussion and Conclusions

TCM's retrieved context is best implemented in a generate recognize model of free recall. Such a model simulated temporal dynamics and total number of words recalled more accurately than a response suppression model, and requires only the addition of an extra context layer and a recognition process to the basic TCM ideas of a model with retrieved context and probabilistic response. Such a model is able to simulate probability of 1st recall, lag-probabilities, temporal dynamics, repetition errors, intrusion errors, and total number of words recalled with only 4 free parameters: a softmax parameter, a contextual change parameter, learning rate, and number of rejections needed to stop recall. Response suppression implementations are more complicated and don't work as well. The best response suppression model obtained in our simulations required the addition of a stopping mechanism, a response suppression mechanism, decay in the

response suppression mechanism, and an as-yet-undiscovered method of modeling temporal dynamics.

Response Suppression Difficulties

It is impossible to say that no potential response suppression model exists that can fit free recall data since new, complex mechanisms can always be added to try and get a model to work. Generate-recognize versions of TCM are to be preferred because they fit the data naturally, and parsimoniously, without the additional need for other mechanisms. A large number of response suppression models were attempted, of which only one example from each class was presented here. However, there were many, many attempts to try and get response suppression to work, all without success. What can be said is that generate-recognize versions of TCM lead to a model that stops recall in human-like fashion and that models temporal dynamics as a natural consequence of its function. Response suppression models create difficulties in simulating this data that are enormous.

The problem that exists with the response suppression models presented here is that too much depends upon item strength. In both versions of the model, recency and lag-recency are effects that were due to differences in item strength. At the beginning of recall, for example, the contextual pattern that was used as a cue by the model was very similar to the contextual patterns associated with items at the end of the simulated list. These items were cued more strongly, and had a greater change of being given as a response. This property of the model thus led to recency results. For response suppression models, temporal dynamics and stopping recall also depend upon item strength in some fashion. It may be practically impossible to get a pattern of activation in the item layer that allows for recency, long-term recency, lag-recency, correct temporal dynamics, and

causes the model to recall a similar number of words as human subjects. On the other hand, in generate-recognize models, the number of unsuccessful generations was used to simulate temporal dynamics and stopping recall. This made it easier to simulate recency and lag-recency results, and it was also in accordance with empirical findings. For example, subjects have a tendency to stop recall after about 10 seconds without a response (for a review of temporal dynamics and free recall, see Wixted & Rohrer, 1994). There is a clear link between temporal dynamics and stopping recall in the memory literature; recall is stopped after a set amount of responseless time. This finding is very similar to the performance of the generate-recognize model, which used word generations as a measure of time, and which always stopped after a given number of unsuccessful generations.

Other Models Including TCM

There are several key differences and similarities of our model to TCM. The model, as mentioned before, uses a TCM-like retrieved context in order to simulate recency and lag-recency effects. There are two key differences, however. The context-item associative strengths are equal to the corresponding item-context associative strengths in the model, whereas in TCM these associative strengths are not symmetric. Also, it is claimed that TCM simulates lag-recency asymmetries due to the process of contextual change. Pre-experimental contexts for a given item are incorporated into the contexts of subsequent items, and not previous items. In the model, this property can also occur. However, lag-recency asymmetries are due to the effects of end-of-list recency and the small number of words recalled in a recall trial (see below). Recency effects

influenced lag-recency asymmetries to a far greater extent than the process of contextual change.

The current model differs from the Davelaar et el. model of recency by having one mechanism of recency effects instead of two. However, the model was meant to simulate recall in the absence of rehearsal or strategy, and so simulates situations where Davelaar's model predicts only one mechanism would be used. The model thus takes no theoretical stand on whether recency is the product of one or two brain mechanisms.

The generate-recognize model is also structurally similar to the Dennis and Humphreys (2001) model of word recognition. Like the Dennis and Humphreys model, memory items are represented by single units, and are associated with distributed contextual patterns. The process of recognition the model performs after generating memory items is functionally similar to that of the Dennis and Humphreys model as well. So although the model was developed independently of the Dennis and Humphreys model, it can be thought of as based on a mixture of TCM (because of its temporally changing context) and Dennis and Humphreys (due to its structure and recognition decision processes).

Multi-trial Free Recall

The model does not accurately simulate multi-trial free recall. As recall trials increase, the number of words recalled does not increase on any trial past trial two. This result would seem to be a weakness of the model, but it is not necessarily the case. The model was meant to simulate free recall in the absence of strategy use and rehearsal. Multi-trial free recall cannot exist without the possibility of rehearsal or strategy use, however. When a subject recalls items, they also have the opportunity to rehearse item or

organize them. This organization or rehearsal may benefit them on subsequent recall trials. Even if subjects perform a task that inhibits frontal lobe involvement during recall, they can still use rudimentary strategies like concentrating on unfamiliar items during subsequent study sessions. So the model simulated free recall only in the absence of strategy use.

Semantic Strategy Version of the Model

The generate-recognize version of the model was expanded in other work (Gilbert & Becker, in preparation) to simulate strategic free recall. This expanded model contained structures that had interesting implications for non-strategic free recall, however. The addition of a semantic layer and a more complex recognition procedure allowed for a more detailed examination of intrusion and repetition errors, and their effect on recency results. This expanded model acts as an existence proof for an entirely new theory of lag-recency.

In TCM, lag-recency results occurred for two reasons. First, memory items presented consecutively or in close proximity during study tended to be recalled consecutively. In TCM, as well as the current memory model, this was due to the fact that they were associated with similar contextual representations, and so retrieving one item caused the contextual cue to be changed to one that was more similar to that of items presented in close proximity to the recalled item. Secondly, items presented after other memory items tend to be recalled better than items presented before. For example, in the memory list PIG, CAT, DOG, presented in that order, DOG would be more likely to be recalled after CAT than PIG would be. In TCM, this is due to words presented after a list word having that list word's pre-study contextual associations associated with it. In the

example, CAT has pre-study contextual patterns that are associated with it, these patterns change the context when CAT is presented during study, and this new context is associated with DOG. PIG is not associated with the pre-list context, and so its context is not as similar to CAT's as DOG's context is. (For a more detailed explanation, see Howard & Kahana, 2001). The expanded model simulates this lag-asymmetry a new way; and this new explanation and model will be described in detail below.

A number of changes were made to the structure and function of the model. Some of these changes are outside the scope of the current work and involve strategic processes, but other changes allow the model to more realistically simulate non-strategic free recall. More specifically, intrusion errors were modeled more realistically, and recognition decisions were made on the basis of more information. When these more realistic kinds of repetition and intrusion decisions were made, models of free recall simulated recency and lag-recency data in an interesting way. This second model (Model B) will be described in the below section. The previous model will be described as Model A.

Not all of the models functionality will outlined in the following description. The model contains parts that were used to simulate strategic free recall; these aspects will be skipped over, and were excluded from the simulations.

There were three significant additions to Model B: a semantic layer, the implementation of more kinds of intrusions, and a more complicated recognition decision procedure. The semantic layer was only used in recognition decisions. Each unit in the semantic layer represented a semantic feature. Each item in the item layer had a semantic pattern associated with it prior to simulation. These patterns were randomly generated,

(by generating a weight matrix between the item layer and semantic layer with random weight values) but with a rule that minimized the amount of overlap between semantic patterns (list words were not from the same category in the simulated human studies). List words could only share a given number of features. Any time a "word" unit was active in the unit layer, it would activate the corresponding pattern in the semantic layer. The activity in the semantic layer would persist across subsequent word presentations during study, or word recalls at test. Activity could either decay over time for each unit, or could change as a result of subsequent word presentations, in a similar fashion to retrieved contexts in the context layer. The decay version will be presented here, because it illustrates several interesting aspects of the model (see below). At each new time step (as measured by the presentation of a word item or an attempted word item generation at test), the current value of each semantic unit was multiplied by the decay factor d, where $0 < d < 1$. Similar results can be simulated with both models, though.

Model B was similar structurally to Model A, with a context layer that changed according to retrieved contexts, a study context layer, an item layer, and the additional semantic layer. It performed identically at study, except for activating units in the semantic layer according to pre-existing associations. No learning occurred between the item layer and semantic layer during study, so this activation had no purpose in non-strategic free recall. During recall, the semantic layer did not cue the item layer. The model can be modified so that this takes place. There is evidence that even in non-strategic free recall, the amount of semantic relatedness between just-recalled word items (as measured by LSA cos θ) and candidate response items influence the probability that a given word will be recalled. In other words, when a certain word is recalled during a free

recall task that controls for strategies, if the next word is related to the just recalled words (i.e. has a similar LSA representation), then it is more likely to be recalled. If the model cues the item layer during recall with both semantic and contextual layer, then these sorts of results can be simulated. It is merely a matter of increasing the amount of semantic layer cuing until

So Model B performed functionally identically to Model A until a candidate word had been generated. Then a more complicated recognition decision was made involving the semantic layer, and the context layer. Comparisons were made between the semantic, and contextual associations of the generated word, and the current levels of activity in these layers. For all comparisons, similarity was measured using the same equation:

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{a}_i}{\|\mathbf{a}\|\|\mathbf{a}_i\|} \tag{13}$$

where $\mathbf{a}$ is the current level of activity in the semantic or context layer, $\mathbf{a}_i$ is the semantic or contextual representation of the current item, and $\cos\theta$ is a measure of the angle between these two vectors. Note that since no learning took place between the item layer and semantic layer in this simulation, all $\mathbf{a}_i$ values for this layer were generated prior to the simulation and put into the equation when needed. $\mathbf{a}_i$ could have been generated from the current pattern of activation in the item layer, and weights between the item layer and the word or semantic layer. Since there would have been no change in these weights because no learning would have taken place, this step was excluded.

Recognition decisions were made using the angle between the two vectors instead of calculating the difference between the two vectors. This change in recognition decision making was necessary because of the function of the semantic layer. In the semantic

layer, activity levels were not constrained in any way. This made the difference between two vectors a poor basis for judging similarity. Comparisons of similarity for vectors that could be of different lengths should be made according to direction. Take, for example, two vectors that will represent the semantic activation of two words: word 1 - [1 0] and word 2 - [0 1]. Let us say that the model gives word 1 as a response, and then generates the word four more times(maybe giving the word as a response, a repetition error, or maybe correctly rejecting it as a repetition, it doesn't matter). After the last generation of word 1, the activity level in semantics is [4 0] without taking decay into account, and, just as an example, [3.5 0] with decay (the decay parameter is around .84 in this example). A comparison between the associated semantic activation of word 1 [1 0] and the current semantic activation [3.5 0] leads to difference of 2.5. Compare this situation to one where the model first gives word 1 as a response and then generates word 2. The activity level in semantics after the generation of word 2 is [1 0] before decay, and around [.84 0] after decay. The difference in activity levels between word 2 [0 1] and the activation in semantics [.84 0] is 1.84. Word 2, with a completely orthogonal semantic pattern, has been judged to be more similar to Word 1 then a semantic context generated purely from Word 1! Comparisons made by the angle between the vectors are much more accurate – then angle between [3.5 0] and [1 0] is 0 degrees, whereas the angle between in [0 1] and [,84 0] is 90 degrees. In this case, the angle correctly represents that word 1 is similar to repeated generations of its own semantic associations, but very different from word 2's semantic associations.

After a candidate item was generated during recall, the comparisons made in Equation 13 were used in a two step process. First, a fluency decision was made. If the

generated word was associated with a semantic pattern that was similar to the current activity in semantics, and was associated with a word form pattern that was different to the current activity in the word form layer, then it was given as output automatically.

$$z_4 > (\cos\theta_{semantics} + \cos\theta_{context}) > z_3 \qquad (14)$$

Note that $\cos\theta_{context}$ involves comparisons in temporal context only, the study context layer was not used in this equation. If a generated word passed the fluency test, it was automatically given as output. Secondly, an overt recognition decision was then made using Equation 8. This two step process was far from the only way possible to make recognition decisions, and prevent repetition errors and intrusions (while allowing them a small percentage of the time). It contains some redundancy, since two separate comparisons were made using the context layer. Recognition was constructed this way to simulate automatic recall influences, and to make them distinct from overt recognition. Equation 14 represents automatic recall, whereas Equation 8 represents an explicit recognition procedure (see below). Equation 8 emphasizes information that was learned by the model during study, namely contextual information concerning the study session, and retrieved contextual information that becomes associated with the study session. Equation 14 emphasizes information that existed prior to study: pre-existing semantic and contextual information, with "priming" of this pre-existing information leading the model to output candidate items without recalling specific contextual information. Future iterations of the model could thus simulate many of the findings of Larry Jacoby and others (see below). For example, the recall performance of seniors could be simulated by decreasing the learning rate of the model, but keeping the thresholds in Equation 8 the same. The model would show similar "automatic" performance, but decreased recall.

The Effect of Repetition and Intrusion Items on Recency and Lag-Recency

When the model was run without any intrusion items (when only list words were part of its lexicon), it did a worse job of modeling recency data. It simulated the recollection of end-of-list items accurately, however, other list items were recalled a significantly lower percentage of the time in comparison to human subjects. This inaccuracy was due to the process of contextual change in the model. The initial context at recall was most similar to the last item on the list. The earlier a word appears in the simulated list, the less similar its context was to this initial contextual activation due to the process of contextual change. It is thus impossible to have a number of words having the same probability of recall with the initial recall cue. However, this is exactly what happens with human subjects. Subjects tend to recall all beginning-of-the-list at the same small, but above zero probability. An easy and realistic way to simulate these small recall probabilities is with intrusion items.

There were four classes of non-list items that were simulated. There were two kinds of intrusion error items: semantic intrusions, and contextual intrusions. Semantic intrusions were defined as those that contained a similar semantic association to the currently active pattern in semantics, and just enough contextual similarity to be given as output by Equation 14. "Contextual" intrusions contained similar contextual patterns to temporal context, and shared some semantics features, enough to be given as output by Equation 14. The semantic layer in the model was strongly involved in semantic intrusions, and the temporal context layer was involved in contextual intrusions. Intrusion errors were also created when intrusion items were created with similar study context patterns. These sorts of items would not exist in free recall trials where one list was used.

However, when subjects study a number of lists of words, and are asked to recall words from one list only, they tend to recall words from other lists as well (Zaromb et al., 2006). These kinds of prior-list intrusions could be simulated using the model; prior list words would have a study context similar to list words.

There has been a lot of work done by Roediger on the properties of a list of words that lead to intrusion errors (Gallo & Roediger, 2002; Roediger & McDermott, 1995; Stadler, Roediger, & McDermott, 1999). Roediger has demonstrated a number of word list factors that significantly influence the probability of intrusion errors. The word lists used in the above simulation were from Howard & Kahana (1999). Neither the intrusion error rates, nor the properties of this word list were known, so no simulation of intrusion errors from this list could be made. It would be problematic to simulate intrusion errors from other free recall lists, since strategic factors could easily change intrusion error rates. An example intrusion error simulation is given in Figure 2.14. Semantic and prior-list intrusion rates can easily be made to increase or decrease to whatever levels are required by simulation. Fitting an exact data point with "contextual intrusions" was much more difficult (see below discussion) In general, intrusion items that are unrelated to list words do not occur frequently enough in free recall to significantly effect the performance of the model in other areas. More interesting, however, was the generation of intrusion items and subsequent rejection by the model. This process played a crucial role in simulating one-trial free recall.

There were two other kinds of non-list items. The first kind was those that were generated but rejected by the model. They had contextual patterns that overlapped with the current context, were generated and then rejected as being not on the memory list

These generated items were very useful to the model in simulating recency. The reason why can be illustrated with a simple example. The model had a tendency to recall only end-of-list items. Those non-list items that had contextual patterns that overlapped with these end-of-list items also tended to be recalled Take, for example, an end of list item with a stored contextual pattern of [.7 .5 .4 - - - - - - - - - -], where "-" equals some non zero, but small value. Non-list item [.5 .4 - - - - - - - .5 .6 - -] is generated, because the contextual pattern that cues it is similar to the end-of-list items' contextual association. This non-list item will be rejected, but it will change the temporal context to a new contextual pattern. This new contextual pattern contains activity that is dissimilar to the end of list item (the .5 .6 at the end) and so will be more similar to items that occur earlier in the list. In generating an intrusion, the model has increased its chances of subsequently generated list items that are not from the end of the list.

An interesting related point is that the model's lag-recency results were related to this phenomenon. At the beginning of recall, the model's contextual activation was similar to end-of-list items in order to simulate recency. If the word at the very end of the list was recalled, there would be no words a positive lag away from that word, and so this recall would not be included in positive lag calculations. If the word second from the end-of-the-list was recalled, the contextual cue for the next recall would change to one that was similar to both the just-recalled item, and the last contextual cue. In this case, the word a +1 lag away would be the overwhelming favourite to be recalled next, since it would contain a temporal context representation similar to both the old context, and the second-to-the-end-of-the-list word's context. Recency thus leads to lag-recency asymmetry in the model (a similar idea was developed independently in Farrell &

Lewandowsky, in press). Because the model started with a contextual cue that was similar to contextual associations of end-of-list items, it tended to recall items at the very end of the list. Whenever it recalled an item that wasn't as close to the end of the list as possible without being a repetition, the next item recalled tended to be a small positive lag away, since the cue being used to generate the item is still partially made up of an end-of-list cue, and words that are a positive lag away are words that are closer to the end of the list. Only in the case of a number of recalls occurring from the beginning or middle of the list does the contextual cue change to one that is not related to end-of-list items.

Another interesting point was that the model could use intrusion errors to simulate an initial first-position primacy effect (as in Howard & Kahana, 1999). If non-list items had contextual associations that were represented by units in the temporal context layer that were not used by retrieved contextual patterns, and if the initial contextual activity prior to study also used these units, then the first word on the list would be recalled with a greater probability. For example, if all pre-existing contexts for list words were of the form [0 0 x y...] and the initial context pattern was of the form [a b c d...], then only associations with this initial contextual pattern would cause a list word to be associated with a pattern that used the first two context units. The first word on the list was always initially associated with this contextual pattern, and so it had contextual associations that used the first two context units comparatively strongly. Subsequent list words had this effect much diluted by the process of contextual change. If non-list items had contextual associations that also used these two initial context units, then any generation of a non-list item would change the current context to a pattern that would favour beginning-of-the-list items more than middle-of-the-list items because of this process. This result

could also be simulated by a process of contextual change between study and recall, if current contextual activity changed in vector dimensions outside of those of the list-item contextual associations.

A final point regarding intrusion errors concerns the difficulty in getting exact intrusion error fits. Non-intrusion, non-list items were easy to create. Word items were given randomly chosen contextual patterns. If first position primacy effects were desired to be simulated, then these contextual patterns sometimes included units outside of list-item's contextual associations. The number of non-list items was chosen according to the computational limits of the computer doing the simulating. The magnitude of these random contextual vectors was then increased or decreased systematically until the data was best fit. It was also not impossible to fit semantic intrusions. Semantic intrusion items were not generated randomly. It was quite reasonable to create non-list items by hand that simulated being in the same category as list words, however. These semantically similar non-list words were given semantic associations very similar to list words, and somewhat similar contextual associations. This "hand-wiring" of semantic intrusion items does conform to empirical result, though (since words of belonging to the same category can often be found in similar contexts – this is the entire basis of Landauer & Dumais's (1997) LSA paper). Prior-list intrusions were also easy to simulate for similar reasons.

There were two difficulties in creating non-semantically related intrusion items. First, these kinds of errors occur a small percentage of the time. In a simulation of 40 subjects who recall an average of 4 or 5 words, only around 200 total words will be recalled during the entire simulation. The model may not recall a single non-semantic

intrusion item, and yet still be accurately simulating human subject performance. To combat this problem, either many more subjects needed to be simulated, or intrusion probabilities needed to be taken directly from the model. A more serious problem was that there was no way of knowing a priori what would be used as a cue at any given point in recall. Both semantics and contextual layers change probabilistically and somewhat unpredictably over time; even if "handwiring" exact contextual associations for intrusion items were attempted, it would be impossible to guess at what contextual associations would overlap with contextual cues at any given point in time during recall. Non-semantic intrusion errors are therefore best modeled using a very large number of non-list items with randomly generated contextual patterns. Non-list items with contextual associations similar to current contextual cues would then exist purely because of chance and large quantities. Computational limits prevented a large number of such non-list items to be simulated, however. A smaller number of non-list vectors was instead created with contextual associations similar to end-of-list items, since it was these items that tended to be recalled. This solution was not considered to be ideal.

Semantics and the Model

The addition of a semantic layer allowed the model to more accurately simulate intrusion errors. However, why go to all the trouble with a new rule in model B and a new method of making recognition decisions? There are two reasons. One interesting feature about how semantics changes over time is that this method of change could potentially be applied to the entire model. Contextual change, instead of occurring due to a TCM-like retrieved context, could instead occur due to this sort of "priming" with gradual decay. There would be no known modeling benefit for constructing the model in

this way. However, this way of viewing contextual change explains away an interesting conceptual problem with retrieved context models like TCM. The problem is this: why would humans have a retrieved context mechanism in their brain? In a model, it help fits the data. In the context of human memory performance, it is less intuitive as to why such a mechanism would exist.

Constructing contextual change in terms of "priming" with decay answers this question in terms of fundamental neuron properties. During study, a pattern of "neuronal activation" exists in the brain representing current contextual information. When a study item is presented or retrieved, associated contextual information causes "neuronal activation" in the same functional brain areas. "Neuronal activation" refers to the rate that a neuron fires action potentials; in rate coded brain models like the ones described here, a particular level of activation in a unit represents this neuronal rate of fire. A neuron does not magically switch from an increased firing rate to a baseline rate as soon as learning occurs. The time it takes a neuron to return to its baseline state causes the activation levels in hypothetical contextual areas to be a combination of retrieved and current contextual patterns. So because retrieved and current contextual information is represented in the same functional brain areas and because neuronal activity is expressed by a rate of neuronal firing that gradually dies down to a baseline level, during a memory study, items are associated with contextual patterns that are a combination of current and retrieved contextual activity. The parameters that determine this rate of decay and activity are based on measurable properties of neurons, whereas with a TCM-like process of contextual change, contextual activation levels are artificially constrained by arbitrary parameter settings.

Constructing a model where this "priming" plus decay process occurs in semantics was a good first step to constructing an entire model based on this process. Preliminary comparisons between a TCM-like process of contextual change and this new method showed that although the two methods required different parameter settings, they could yield the same intrusion error results for at least one data set, and they appear to work in the same fashion. A potential future direction of the model would be to construct the model using this kind of change in both the contextual and semantic layers.

A second reason for a different method of semantic activation change involves another potential future direction of the model. It is not necessarily the case that free recall always involves a generate phase and a recognize phase. There has been a lot of memory work on implicit memory influences. Larry Jacoby, for example, hypothesizes that memory tests tap into both recollective (generate-recognize like) processes as well as implicit (automatic) processes (an example of these ideas and the process-dissociation procedure used to test them can be found in Jacoby, Toth, & Yonelinas, 1993). The measure of the angle between the current semantic and word form activity, and semantic and word form associations of the generated memory item can be viewed as a very rudimentary simulation of a measure of fluency. Other parts of the brain such as areas that process word form and word sounds could also be simulated in this fashion. Whenever a word was generated by the model, similarity to semantics (and word form and sound in a hypothetical expanded model) could cause a word to be given as output without a contextual comparison. So a semantic layer with a similarity comparison based on vector angles could be conceptualized as part of an implicit memory system. Future work on the model could focus on this implicit/recognition difference.

Ideally, both of these ideas could be combined into a single model. The model would have an episodic component, where learning takes place, equivalent to the study context layer in the current model. The model would also have non-episodic components, including semantics, where presentations or generations of items would result in heightened activity ("priming") which would gradually decay, and which would act in a similar fashion to retrieved context. Both episodic and non-episodic layers could be used to cue memory. A PFC layer could control the amount of contribution each layer made as a cue, and could control how implicit or explicit each layers contribution was. For example, when adopting a semantic strategy, the PFC layer could extract semantic cues from a list of words, and cause them to be used as cues during recall. This PFC function would change the semantic layer's contribution to recall from an implicit cue due to residual activity, to an overt, explicit recall cue. This kind of model is the end goal of the current modeling work

A Lateral Inhibitory Model Of Retrieval Induced Forgetting – Chapter 3

Modern theories of memory suggest that people forget memories because other memories compete with them (e.g. Mensink & Raajimakers, 1988). Recently, retrieval induced forgetting studies have suggested that an executive function is sometimes necessary to overcome this competition (Anderson, 2003). In this view, in order to recall a memory item, memory items more strongly associated with recall cues must be inhibited by an executive process. In this paper, we present an explanation of retrieval induced forgetting that does not require such a process. We present a model able to account for a wide range of both free and cued recall data, as well as retrieval induced forgetting studies that previous models could not account for.

There are a number of studies that have demonstrated that recalling an item in a cued-recall task can reduce the effectiveness of subsequent recalls (Smith, 1971, 1973, Tulving & Arbuckle, 1963, Roediger & Schmidt, 1980). Recently, this kind of effect, called retrieval induced forgetting (RIF), has become synonymous with the retrieval practice paradigm developed by Anderson, Bjork, & Bjork (1994). Several competing explanations of RIF have been put forward. Perhaps most prominently, Anderson and his colleagues claim to have shown the superiority of an item inhibition account over other accounts of RIF (e.g. Anderson & McCulloch, 1999; Anderson & Speelman, 1995, Anderson, Green, & McCullock, 2000). Although the idea of inhibition as item suppression is controversial (MacLeod, Dodd, Sheard, Wilson & Bib, 2003), Anderson's line of research represents some of the idea's greatest support.

In the Anderson retrieval practice paradigm, category/semantic exemplar word pairs are first studied. Next, half of the studied items from half the studied categories are

practiced before recall, often with a cue plus word-stem completion task (e.g. DOG – HO___ is practice for DOG-HOUND). Then, after a delay period, all the word pairs are tested using the semantic category as a cue. During testing, there are thus three different kinds of items: practiced (RP+), related but unpracticed (RP-), and unrelated, unpracticed items (Nrp). For example, if only DOG-HOUND had been practiced after study, DOG-HOUND would be an RP+ item, DOG – BEAGLE would be an RP- item, and METAL-SILVER would be an Nrp item. The common finding is that RP- items are recalled a lower percentage of the time than are Nrp items, with RP+ items being recalled the highest percentage of the time (Anderson et al., 1994). Thus, retrieving an item in this paradigm lowers the probability of retrieving related items. Anderson and his colleagues have used this method to build the case that RIF is caused by item inhibition. According to Anderson, in order for an item to be recalled, potentially interfering related items must be suppressed. When an RP+ item is practiced, it leads to the inhibition of competing items so that when the practiced item is cued again, it can be recalled more easily. This postulated inhibition is direct, and occurs during study at the level of item representation; consequently, the inhibited items will be difficult to retrieve by any cueing method. Also, items highly associated with the category cue will undergo the greatest inhibition, and so item inhibition is proportional to categorical association (Anderson, 2003).

There are two crucial aspects to Anderson's account of RIF. First, if an item is suppressed, it should be difficult to access using any cue, and not just using the semantic category it was paired with during study. Anderson refers to this property as the *cue-independence* effect (Anderson, 2003). Support for this aspect of Anderson's account of RIF has been mixed. When new cues used at test are related in some way to cues used at

study and practice, then RIF is found. Anderson et al. (2000) used extra-list items during testing that were semantically related to list items. For example, a RIF procedure with items like METAL-SILVER and METAL-GOLD might have a testing phase with the cue "tell me a word related to jewelry that begins with an S". There are a significant number of additional studies that use new, list-related cues at testing (Carter, 2004; Johnston & Anderson, 2004; Shivde & Anderson, 2001; Veling & van Knippenberg, 2004). All of these studies do find RIF effects consistent with Anderson's cue-independent property of item suppression. However, none of these studies use cues that are truly independent in the sense of having no prior association to the studied word pairs.

Perfect, Stark, Tree, Moulin, Ahmed, & Hutter (2004) examined whether RIF would take place when completely unrelated cues were used at testing. In Experiment 3, for example, there were two study phases. The first phase used word pairs containing category exemplars and completely unrelated words. Subjects were presented these word pairs in sequence until they achieved a 60% success rate. The second study phase involved the usual presentation of exemplar-category word pairs. So the word pairs DRESS-HAMMER and SPIDER-PLIERS might first be learnt, and then TOOL-HAMMER and TOOL-PLIERS might be presented in the second study phase. After a practice phase for half the word pairs in half the categories, testing was done either with the categories, or with the unrelated words paired with the exemplars at study as a cue. According to Anderson's cue independence property, if a pair like TOOL-HAMMER was practiced, recall of PLIERS should be impaired, regardless of whether TOOL or SPIDER was used as a cue. However, Perfect et al. found RIF effects only when using the categories as cues, but no RIF whatsoever when using truly independent cues.

The support for Anderson's *cue-independence* property is thus mixed. Unless some fault is found in the Perfect et al. study, a general inhibitory account of response suppression does not explain why word pairs tested with unrelated cues fail to show any RIF. However, RIF results from studies using related cues at test are difficult to explain without invoking an item level inhibition explanation. An accurate account of RIF must explain why a wide variety of related cues lead to RIF, but unrelated cues do not.

A second important aspect of RIF studies is that RIF depends on the pre-existing strength of association between category and exemplar items. If word pairs with weak category-exemplar associations are used, little or no RIF effect is found (Anderson, Bjork, & Bjork, 1994). For example, weak semantic associations like WEAPON-NAIL and WEAPON-FOOT lead to no significant RIF effect, whereas word pairs like WEAPON-SWORD and WEAPON-PISTOL lead to RIF. Anderson (2003) refers to this property of RIF as *competition-dependence.* Items that are highly related to a category cue are somewhat activated during practice, according to Anderson, and are inhibited to the degree that they were active. In the above example, if WEAPON-SWORD, WEAPON-FOOT and WEAPON-PISTOL were all studied, and WEAPON-SW___ was given as practice, there would be activation of SWORD, but also some activation of PISTOL, due to its strong association with WEAPON. There would be little or no activation of FOOT. During the practice session, some mechanism would "punish" both PISTOL and FOOT by inhibition, according to the amount they had been erroneously activated during the practice session (little for FOOT, and much more comparatively for PISTOL). Then, at testing, this inhibition would linger, and make the difficulty of retrieving these words commensurate with the prior inhibition.

Anderson, Bjork, & Bjork (1994) also ran an experiment where both weak and strong category exemplars were intermixed. All combinations of strong and weak exemplars were used: strong exemplar RP+ and RP- items (SS), strong exemplar RP+ items and weak exemplar RP- items (SW), weak RP+ items and strong RP- items (WS), and weak exemplar RP+ and RP- items (WW). So a SS condition might have word pairs like WEAPON-SWORD and WEAPON-PISTOL, and a WW condition might have WEAPON-FOOT and WEAPON-NAIL. The SW and WS would intermix these pairs (for example WEAPON-SWORD and WEAPON-FOOT) with the SW condition having the strong exemplar practiced (WEAPON-SW___), and the WS condition having the weak exemplar practiced (WEAPON-FO__). Anderson et al. found no retrieval induced forgetting in the WW and SW conditions, but did find RIF in the WS and SS conditions.

These results supported the *competition-dependence* property of Anderson's item suppression account of RIF. Strong items are more likely to be given as a response using a category cue, and thus must be more strongly inhibited when learning a different category-exemplar pair. Thus, the practice phase in a RIF study inhibits strongly related pairs more than weakly related pairs, and this leads to RIF in conditions where strongly related items are unpracticed (the WS and SS conditions). A subject studies these strong items at study, inhibits them to a large extent at practice, and so recalls them less during testing. Weak related items require little or no inhibition during practice, so there is little or no RIF where weakly related items are unpracticed (the WW and SW conditions).

While Anderson's item-based inhibition theory of RIF explains a wide range of data, it leaves a number of issues unresolved. First, Williams and Zacks (2001) failed to replicate both cue-independence, and competition dependence results. They did,

however, find a trend of strong exemplars leading to more RIF than weak exemplars, although that trend was non-significant (they didn't test intermixed groups). Other studies, such as Bauml (1998), have found support for competition dependence in RIF. With only a single conflicting result, the competition dependence property is fairly well established. However, the cue-independence property is less firmly established, considering that no study has found RIF with truly independent cues.

Computational models can help to further elucidate the mechanisms underlying RIF. A crucial part of Anderson et al.'s (1994) item suppression account was a proposed ratio-rule model, which consisted of categories, exemplars, and strengths of association between the two. Only semantic associations were modeled, and retrieval was a function of association strength. Anderson et al.'s ratio-rule model of these results was somewhat ad-hoc, as it relied upon strong exemplars undergoing more learning during practice than weak exemplars. Learning rate tends to be a negatively accelerated function of prior strength, as Anderson et al. (1994) themselves noted in Appendix A. On the other hand, a more recent model of RIF by Norman et al, also using item inhibition does accurately model these results (Norman, Newman, & Detre, 2007, see Discussion for details). Finally, regardless of whether item inhibition accounts can explain the full range of results, no other explanation has been put forward to date that explains crucial RIF properties and does not include item inhibition. Lateral inhibitory models, (Anderson, Bjork, & Bjork, 1994), for example, show RIF effects for SW subjects and no RIF for WS subjects, the exact opposite of the behaviour of human subjects

In endorsing an item suppression account on the basis of the above studies, Anderson (2003) rejected three competing explanations. All three classes of competing

theories are described in terms of retrieval practice studies involving category-exemplar pairs: 1) Lateral inhibition involves strengthening inhibition between exemplars when the association between one exemplar and the category is also strengthened. 2) Associative unlearning involves a weakening of the association between an exemplar and its category, when another exemplar-category pair is presented. 3) Finally, response competition theories posit that RIF effects are caused by the strengthening of an exemplar-category association when they are paired up, resulting in greater response competition at retrieval. So if WEAPON-SWORD is practiced, SWORD will be more strongly associated with WEAPON, and when WEAPON is given as a retrieval cue, SWORD is more likely to win a competition between other highly associated items like PISTOL. Both associative unlearning and response competition theories have had difficulties explaining the *competition-dependence* and *cue-independence* properties of RIF. It is the purpose of this paper to describe how a neural network model employing response competition in the form of lateral inhibition can account for both of these properties.

Lateral inhibition was rejected as an explanation of RIF findings because of its inability to simulate competition dependence properties of RIF. In Anderson (1994), both weak and strong semantic exemplars were used in the study. Lateral inhibition predicts that strong exemplars should inhibit other items more than weak exemplars. So practicing strong exemplars should lead to a significant decrease in performance for related weak exemplars in a RIF study. However, practicing strong exemplars was shown to have no significant effect on weak exemplar recall.

An assumption of this explanation is that strong and weak exemplars are represented in the same brain area, and are recalled by the same cues. In Anderson (1994), both kinds of exemplars are represented in a single semantic layer. However, both semantic and episodic (contextual) cues might be used in recall, with exemplars having both semantic and episodic representations. If this is the case, a lateral inhibitory/ response competition model would demonstrate different results.

This paper describes an attempt to model the above RIF results. The goal of the paper is to demonstrate that a response competition theory can accurately simulate both *competition-dependence* and *cue-independence* properties of RIF, and to show that item inhibition is not the only theory that can account for RIF results.

Model And Simulations

The RIF task was simulated in a neural network model designed to emulate human performance on free recall tasks. A schematic of the model architecture is shown in Figure 3.1. The model consisted of four of layers, including the item layer, word form layer, semantic layer, and context layers. The item layer represented the to-be-remembered items in the RIF task; a single unit in this layer was representative of a single item. The semantic layer represented the meaning of items presented during a RIF task, and also used localist representations. This very simplistic localist representation of semantics could be thought of as representing semantic categories. The model works equally well when a distributed representation of semantic features was used (and maybe better, see Appendix D) .There was also a layer representing word form (for the experiments simulated here, this layer represents word phonology). Here, words were represented by a distributed pattern of activation with each unit in the layer having an

activation value between 0 and 1. The context layer consisted of two separate sub-layers. The first contextual sub-layer represented things that stayed relatively constant during a RIF task, such as surroundings and task instructions. The second sub-layer changed across time according to retrievals. This part of the model was originally developed to simulate human performance in a variety of free recall tasks (Gilbert and Becker, in preparation). Note that this time-varying part of the context layer (the second sub-layer) was not required for simulating performance on RIF tasks, and so will not be discussed further in this paper. Both context sub-layers used distributed representations.

All layers were fully interconnected with the item layer. Weights existing between units in different layers were updated according to a simple Hebbian learning rule. During all phases of the simulated RIF experiments, any time there were concurrent patterns of activation in connected layers, learning took place between the layers according to the following Hebbian learning rule:

$$\mathbf{W}_t = l(\mathbf{x}\mathbf{i}^T) + \mathbf{W}_{t-1} \qquad (1)$$

where **W** is the weight matrix between a second layer (context, semantics, or word form) and item layer, l is the learning rate, **x** is the activation in the second layer, t is the current time step and **i** is the transpose of the item layer activation.

Prior to the simulation of a task, the model's prior knowledge was represented in its pre-existing weight values. First of all, word items were strongly associated with their category by randomly setting weights between these two layers to values between .8 and 1 for connections between word nodes and their corresponding category nodes. For example, the weight between BANANA and FRUIT might be given a value of .87. Non-list word items were also associated with categories in a similar way. Word items

were also weakly associated with a random pattern of activation in the contextual layers, representing all the contexts in which this item has been seen previously. To simulate this association, weights between these two layers were randomly set between values of 0 and .2. In the word form layer, patterns were pre-defined to represent forms associated with the items. Weights were set at values such that activating an item unit would reinstantiate these patterns. To do this, Equation 1 was applied to each word and corresponding phonology pattern, with a learning rate of 1, and with the pattern of activation in the word layer as the output layer and the pattern of activation in the phonology layer as the input layer.

$$\mathbf{W} = (\mathbf{O}\mathbf{I}^{\mathbf{T}}) \tag{2}$$

(where **O** is the set of phonology patterns the items should generate, $\mathbf{I}^{\mathbf{T}}$ is the set of list items, and **W** is the weight matrix between the two layers)

During a standard RIF task, the following occurred. First, the study phase, in which category-exemplar word pairs are presented, was simulated. Each presentation of a category-exemplar pair (e.g. FRUIT-ORANGE) was simulated by clamping activation (at a magnitude of 1) to the corresponding semantic and word layer units. Hebbian learning then occurred as in Equation 1 between the semantic and word layers. At the beginning of study, a random pattern of activity was instantiated in the context layer, where each unit in the context layer was randomly assigned a value between 0 and 1. This represented things that stayed constant during task performance, like surroundings and task instructions. Hebbian learning also took place between these two layers as in Equation 1.

Next, the practice phase, the task of category-cued completion of exemplars (e.g. FRUIT-OR_____) was simulated (for the control conditions, this phase was skipped). For

each practice pair, activation was clamped at the corresponding semantic units, as well as in the word form layer, representing the category and part of the exemplar (e.g. the letters FRUIT-OR). The word form pattern of activation was completed a certain amount of the time, based on empirical findings (Anderson & Bjork, 1994). For standard RIF items, subjects were able to successfully perform the practice phase (for example, given FRUIT-OR, subjects were able to complete it as FRUIT-ORANGE) 81% of the time. In these cases, the full word form pattern representing the category and exemplar was instantiated, and this pattern cued the item layer so that the exemplar was active there (due to prior learning). Then Hebbian learning took place between the word form and item layers as in Equation 1. In cases where the pattern was not completed, no learning occurred. Hebbian learning also took place between the context layer, which contained a pattern of activation that was the same as the pattern of activation during study, and word and word form layers. Note that the learning in the model here was hypothesized to be at the level of word form as well as semantics; since knowledge of orthography was required to perform the practice task, learning was concentrated in this domain to some degree

Testing, where only the category is given as a retrieval cue, was then simulated. For each category cue, an activation of 1 was clamped to the corresponding unit in the semantic layer and a pattern in the word form layer representing the word form of the category was also clamped. The context layer still contains the same pattern of activation that was assigned at study. These layers then cue the unit layer in the following manner:

$$\mathbf{i}_s = \mathbf{W}_s^T \mathbf{s} \tag{3}$$

$$\mathbf{i}_p = \mathbf{W}_p^T \mathbf{p} \tag{4}$$

$$\mathbf{i}_c = \mathbf{W}_c^T \mathbf{c} \tag{5}$$

$$\mathbf{i} = \mathbf{i}_s + \mathbf{i}_p + \mathbf{i}_c \quad (6)$$

(Where **i** is pattern of activation in the word layer, $\mathbf{i}_p$ is the activation in the word layer due to word form, $\mathbf{i}_s$ is the activation in the word layer due to semantics, $\mathbf{i}_c$ is the activation in the word layer due to context, **s** is the pattern of activation in the semantic layer, **c** is the pattern of activation in the semantic layer, **p** is the pattern of activation in the word form layer, $\mathbf{w}_p$ represents the weights between the word layer and the word form layer, $\mathbf{w}_c$ represents the weights between the word layer and the context layer, and $\mathbf{w}_s$ represents the weights between the word layer and the semantic layer)

Once the item layer was cued, one unit in the item layer won the competition and was chosen as the "winner" by softmax competition (Brindle, 1990).

$$p_i = \frac{e^{i\mu}}{\sum_{j=1}^{n} e^{j\mu}} \quad (7)$$

(where p is the probability of item i being selected as the "winner", j is the index of an item in the item layer, and n is the number of items in the item layer, and μ is the softmax parameter.) This simulates a form of lateral inhibition in which the inhibition is divisive (shunting) and the unit with the least inhibition has the highest probability of being activated.

The selected item then underwent a recognition process, the purpose of which was to reject both repetition and intrusion errors. During this process, item activation propagated backwards into the semantic and context layers through the corresponding weights. If the semantic unit corresponding to the correct category was sufficiently active, and the context pattern was different enough from the current pattern, then the

item was given as a response. For both the semantic and context layers, the current level of activity was compared to the activity generated by the item layer:

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{i}_a}{\|\mathbf{a}\|\|\mathbf{i}_a\|} \tag{8}$$

(where **a** is the level of activity in the semantic or context layer, $\mathbf{i}_a$ is the association the currently generated item has in this layer, and $\cos\theta$ is a measure of the angle between these two vectors)

The generated word was said to have been recalled if

$$z_1 < total\cos\theta < z_2 \tag{9}$$

(where z_1 and z_2 were upper and lower thresholds for rejection, and total $\cos\theta$ was the sum of all $\cos\theta$ comparisons done using Equation 8 for semantic, context, and word form layers). This equation allowed the model to reject candidate items if they are too similar (i.e. a repetition of the item just recalled) or too dissimilar (intrusion errors) to the cue used to generate them, so that immediate repetitions were prevented, as well as words judged to be too dissimilar to the cue used to generate them.

Simulation 1: RIF Main Effect

We first simulated the main RIF effect using highly related associates at study and practice. Results are given in Figure 3.2. Practiced items are recalled more than unpracticed items, and, more importantly, unpracticed items in practiced categories are remembered less than control unpracticed items. RIF effects are found in this model due to the fact that practice causes weight values to increase between semantic and phonological patterns and some of the word units. These units are then activated to a greater degree during testing, and softmax competition has them more likely to be given

as a response and competitors less likely to be given as a response. This RIF main effect is very robust, as it is found across a wide range of parameter settings.

Simulation 2: Competition dependence

Studies have been run that vary the degree to which exemplars are related to the category. Most RIF studies use only highly related exemplars, but Anderson, Bjork, & Bjork (1994) used both low and high related exemplars. As mentioned above, studies with all low related exemplars show no RIF at test, studies with both high and low exemplars show RIF if the low exemplars are the ones practiced, and no RIF if the high exemplars are the ones practiced (for examples of each of these cases, see Appendix A). These results have been cited in support of an item suppression account.

The model simulated these findings by assuming several key differences between high and low exemplars. First, low related exemplars had much smaller pre-study weight values between an item and its category. High-related exemplars are almost certainly in some way organized according to category in semantics, and so have greater weight values. Although the link between low-related exemplars and their related category can be either easily recognized or constructed, it is not necessarily the case that this relation is explicitly stored in long term semantic memory in some way (semantic categorization might also occur due to contextual links and active, on-the-fly categorization). The two kinds of exemplar-category pairs may be processed in fundamentally different ways, thus, in our model, low-related exemplars had near zero weight values of zero between word item and category. Secondly, because of the first assumption of a fundamental difference between high related and low related category-exemplar pairs, there were differences in the model in learning the pairs during study, and in trying to retrieve the

exemplars during testing, as will be seen below. High related exemplars had lower learning rates, since they have greater prior associations, and, as cited above, learning rates tend to be a negatively accelerated function of prior strength.

The model simulated all four relatedness exemplar groups: SS, SW, WS, and WW. Simulation occurred in an identical fashion to the purely SS run of the model in Simulation 1, with a few key differences. First, weakly related exemplars had no initial association with semantics, thus, the weights between weak exemplars in the word layer, and categories in the semantic layer were not greater than .1. Secondly, during study, weights between presented word items representing strongly related exemplars and their associated category were incremented only by a small amount. Hebbian learning as expressed in Equation 1 was still used, but with a smaller learning rate. Since the link between a strong related exemplar and its category was very small, it is hypothesized that this link can not be learned to a great extent, and only priming occurs. Learning between context and word layers is small, because preexisting semantic relationships are highly familiar and don't induce a great amount of learning. Weak associates are unfamiliar, and generate a significant amount of learning in both semantic and contextual (episodic) associations.

In summary: high related exemplars were processed in terms of their semantic relation to the associated category during study, which lead to a small amount of semantic learning. Low related exemplars were treated as novel, with little or no permanent semantic association between the low related exemplar and the category in memory. Hence, learning was concentrated in an episodic domain (context) during study. For both high and low related exemplars, significant word form learning took place as

well during practice due to the demands of the practice task. High concentration of learning is simulated by a larger learning rate in the model for Equation 1, low concentration is simulated by a smaller learning rate, and semantic associations have a ceiling associated with them, preventing the learning of an association between category and exemplar greater than 1 (see Table 3.1).

Although simulating practice was the same as normal simulations run above, testing was also simulated differently. Here, cuing was a two step process. First, the semantic layer and phonological layer cued the word layer directly, as in Equations 3 and 4. The softmax function then selected an item probabilistically from the sum of Equations 3 and 4, as in Equation 7 and a recognition decision was made on the basis of Equations 8 and 9. If this generated word item was rejected, a second, longer process was hypothesized to occur. Here the context layer alone cued the semantic layer, as in Equation 5. The softmax function then selected an item from this pattern of activation, and a recognition decision was again made on the basis of Equations 8 and 9.

This model was able to simulate the complete range of relatedness findings. The results of the model are shown in Figure 3.4, and can be compared to Anderson, Bjork, and Bjork's results in Figure 3.3. In both figures, Control A was the control group for exemplars that had undergone retrieval practice, and Control B was the control groups for exemplars that did not undergo retrieval practice. Control items had the same amount of semantic relatedness to their category as their associated experimental group (so in the SW condition, Control A items were strongly related, and Control B items were weakly related to their associated category). There were two important aspects to these results. First of all, all practiced items should be recalled a higher percentage of the time than

their associated Control A counterparts. This aspect of the Anderson, Bjork, and Bjork results was simulated by the model trivially. Secondly, unpracticed items should be recalled a smaller percentage of the time in comparison to their associated Control B counterparts for the SS and WS conditions, but not the SW and WW conditions. Again, as can be seen in Figure 3.4, the model successfully simulated this result.

Practiced words always were recalled better than unpracticed and control words because they underwent additional word form (phonological) learning during the practice stage which increased their chance of being given as a response during test. Strongly related items that were unpracticed are recalled less than control items due to lateral inhibition simulated by the competitive softmax activation function. Phonological learning at practice causes the practiced items to be activated more strongly at test, and to inhibit unpracticed items to a greater extent. Weakly related, unpracticed items are not recalled less than control items, because recall of these items depended mostly upon contextual factors. Contextual cuing was completely independent of the other kinds of cuing, and so phonological learning at practice did not inhibit cuing due to context. For these weak related, unpracticed items, the model often generated an unsatisfactory candidate word using word form and semantics, correctly rejected this candidate word, and then generated a correct response using context. If time was measured in candidate-item generations, the model predicted that responses of weak related items should take longer than responses of strongly related items.

A simplified, general version of the model demonstrated that these effects will be independent of parameter settings. Using a model with the same architecture as above, the model was simplified with the following properties: 1.) During study high-related

exemplars learned in semantics only, low-related exemplars learned in context only. 2.) During practice, both sets of exemplars had only word form learning. 3.) During test, the model first used word form and semantics to cue as many memory items as possible, then it used context. This simplified model was equivalent to two separate completely independent networks, one which involved semantics and word form, the other which involved context. Unpracticed, low-related exemplars relied only on contextual learning, and so no amount of practice changed the probability they were recalled, since practice influences a "separate" network. Practicing low related exemplars gave them word form associations and increased the chance they would be recalled since they could then be recalled with both the semantic/word form network, and the context network. High related exemplars were only ever recalled using semantic/word form associations that exist in the corresponding network, and so practice influences recall probabilities in the usual way. Competition dependence in the model thus depended on the degree that contextual and semantic associations were independent; if these associations were stored in separate areas, and if contextual and semantic cues generated candidate responses at test at different times, competition dependence always occurred.

Simulation 3: Cue independence

There are several different kinds of cue-independence studies (see Appendix B). Anderson, Green & McCulloch (2000) report an experiment that was identical to a standard RIF paradigm, except testing was done using a different cue. For example, a standard RIF study might have word pairs RED-BRICK and RED-TOMATO, might have the practice item RED-BR__, and the test cue RED. In Anderson et al's modified version, study and practice would be the same, but the test cue might be FOOD. Even with the test

cue FOOD, RIF is still found, and this has been taken as support for an item inhibition account. It is normally assumed that people use the cue FOOD to recall study items. However, in a generate/recognize framework, a subject has two choices. They can either use FOOD to cue items, and check them to make sure they were previously studied using previous contextual and category cues, or they can use previous contextual and category cues to generate items, and check them to make sure they are FOOD. Since the FOOD-TOMATO association has not been primed, and there are a huge number of items potentially associated with FOOD, it might be time-consuming to generate items using this cue and verify they were study items. It is much easier and systematic to use recent category cues whose cue-exemplar associations have just been primed (e.g. RED) and then check the generated items to make sure they are FOOD. Subjects know that there are only a very small number of RED things associated with the study they are performing, and they have just had significant practice learning the associations between RED and some of its exemplars. Subjects are thus able to use RED as a cue because they are aware of the link between words cued by RED and words cued by FOOD. Since subjects are presumed to use the normal category cues instead of the cues provided at testing, RIF occurs as in any normal study. The only difference is at the test phase – here subjects use practiced semantic retrieval cues (in this example, RED) to retrieve items, and make a recognition decision on the basis of the semantic category FOOD. (For a similar argument, see Perfect et al., 2004).

The procedure employed by Perfect et al. (2004) involved a second study phase (see Appendix B), where exemplars were associated with unrelated words. Subjects during study learned both category-exemplar pairs (like FRUIT-APPLE) and the same

exemplars were paired with unrelated words (GOLF - APPLE) (see Appendix B for details). Perfect et al. found RIF when exemplars were cued with categories at testing, but no RIF when exemplars where cued with the unrelated words. (see Figure 3.5, ignore MEpisodic and MCategory bars)

The model simulated these results similarly to the above RIF studies with one difference. The unrelated words were given their own unit in the semantic layer, similar to categories, as well as their own word form representation. The second study phase was simulated by presenting the unrelated word together with its exemplar, and having Hebbian learning take place between the various layers. More specifically, the learning that took place was similar to that of the low related exemplars in the competition dependence simulation: there was normal semantic learning, and a significant concentration on contextual learning. Similar to the Perfect et al. results, no RIF was found with the unrelated words (see Figure 3.6). This result was straightforward for the model, as unique cues cause no significant competition in the item layer. Each unrelated word was associated strongly only with one exemplar during study, and so there was very little response competition (the unrelated words are assumed to have some pre-existing semantic associations with other words). This result is only difficult to explain for item suppression models.

Simulation 4: Full Practice

In a full practice condition, the practice phase consists of the full word instead of part of the word (WEAPON-SWORD instead of WEAPON-SW, for example). Anderson & Shivide (in preparation) have shown that full word practice does not lead to RIF, only partial word practice. An assumption underlying the model was that with this full

practice, no learning took place between the word and word form layers (there is no concentrated orthographic learning as there is when a word must be completed, since reading a word is a relatively automatic process).

RIF involving full word learning was modeled identically to partial word learning, except during practice. At practice, since word stems do not need to be completed, there was learning 100% of the time. (In a normal RIF study, subjects do not always successfully complete practiced items like FRUIT-or____. However, with full word learning, subjects merely rehearsed FRUIT-orange.) Also, learning only took place between the word layer and the semantic layer (as in Equation 1), for reasons described in the above paragraph. Since weight values between the item and semantic layers were already very high, and since the weights either get normalized (or the learning rate was smaller in the semantic layer), the amount of learning that occurs at practice was miniscule, and a significant RIF effect did not occur (see Figure 3.7). Due to the smaller amount of learning that took place, there may also be a smaller increase in recall for the practiced words in the full practice condition than in the partial word practice condition, but this effect can be somewhat negated by the 100% chance of learning in the full practice condition compared to 81% in the partial practice condition. Whether or not there is a significant difference in recall between practiced words in the full word condition and the partial word condition depends significantly upon the parameters that are used.

Discussion

The above simulations demonstrate the viability of a lateral inhibitory model of retrieval induced forgetting. Using a generate-recognize model that employed softmax

lateral inhibition, we have simulated competition dependence, cue independence, and several other major patterns of RIF results. Previous models and theories of RIF center on a mechanism of item-specific inhibition; prior work was either based around this view, or was done to examine the limitations of this view, or to criticize it in some fashion. These simulations act as an existence proof of an alternative view, and they show that a simple lateral inhibitory model can simulate results previously felt to be problematic to theories that did not include item-specific inhibition.

The simulation results above also examined the RIF procedure in some detail. Earlier papers on RIF focused only on semantics (Anderson & Bjork, 1994), although later versions of Anderson's theory include a significant focus on the PFC (Anderson, 2003). This paper demonstrates how psychological domains other than semantics could hypothetically influence RIF results. In the remainder of this paper we discuss how the model relates to memory findings and issues, and compare this model to one other computer simulation of RIF (Norman et al, 2007)

Memory

The above simulation can be described well in terms of semantic and episodic memory processes. The model is based on four major assumptions within these domains. 1.) The presentation of a memory item such as a category-exemplar pair leads to an episodic memory. This was represented in the model by the learning that occurs between the item layer and the context layer when category-exemplar pairs were presented, and when practice trials occurred. 2.) If the presented stimulus taps upon some sort of pre-existing knowledge, then this knowledge is activated, and is more available for recollection a significant period of time afterwards. This was simulated by the learning

that occurs between the semantic and item layers during category-exemplar presentation. This learning imposes a ceiling on association strength, as it is a negatively accelerated function of association strength. Pre-existing semantic knowledge can be "primed", but since strong semantic relationships are by definition already well learned, new semantic learning has a small effect. Category-exemplar pairs that do not have a pre-existing (semantic) association do not get primed. 3.) Pre-existing knowledge "overshadows" episodic learning, whereas novel stimuli evoke strong episodic learning. Thus if a category-exemplar pair constitutes a well-learned semantic association, then the association of these two words in a RIF context will be poorly learned. 4.) During recall, pre-existing semantic learning and new contextual learning do not compete simultaneously. This property is manifested in the model by having the item layer not receive input from both semantics and context simultaneously (the item layer is cued with semantics, and then later context).

One way of conceptualizing this fourth assumption of the model is in terms of controlled and automatic processing (as in Jacoby, Toth, & Yonelinas, 1993, see also Jacoby, 1991). During retrieval, semantic and word-form knowledge work in an automatic fashion to influence a subject toward responding with exemplars that have been "primed." Long term semantic associations are recalled and perhaps orthographic knowledge is recalled as well, since knowledge of how a word is spelled is necessary to successfully complete practice trials, and this knowledge could be residually active in some manner during recall. New (contextual) knowledge is remembered by controlled processing during recall. A subject is hypothesized to use contextual/episodic cues to generate candidate recall candidates, and this takes a greater amount of time. Note that it

is not necessary for semantics and context processes to be independent of one another for the model to work (as in Jacoby, 1991); it is only necessary that the strength of a particular semantic association not directly inhibit contextual learning during recall. Also note that semantic influences do not necessarily have to be entirely automatic for the model to work, it is only necessary that if there is "controlled" semantic processing during recall, it should occur to the same degree for practiced and unpracticed items.

The evidence for these four assumptions is fairly strong. As discussed above, there have been groundbreaking papers from Jacoby (e.g. Jacoby, 1991) examining the separate effects of controlled and automatic influences on recall, and so separating these processes in a model of memory is not unusual. However, due to Anderson and colleagues' dominant theory of RIF in terms of item inhibition, RIF has not been described in these terms (although see Perfect et al., 2004). The fourth assumption therefore remains untested. The other three assumptions have been well demonstrated in the memory literature. The evidence that specific episodes can be remembered is vast, as any study of episodic memory supports this first assumption. The chance of a word being recalled is a negatively accelerating function of practice, and so the second assumption stands on solid ground (Hull, 1943 is a classic paper on this topic, and there are many other memory studies demonstrating this property). Finally, there have been a number of studies where pre-existing semantic knowledge has interfered with new learning (e.g. Bukach, Bub, Masson, & Lindsay, 2004).

Semantics, an Aside

Not all category-exemplar pairs used in RIF studies are necessarily associated strongly together in semantic memory. There are at least three potential category-

exemplar relationships. First of all, an exemplar may be strongly associated with a category. Such a pair represents a strong association between an exemplar and a category in some fashion, in some form of semantics or semantic memory. Secondly, an exemplar may be easily identified as belonging to a category when paired with the category name, but may not be associated with the category in semantics. FOOT may be easily identified as a weapon when presented with WEAPON in the context of a RIF study, but FOOT may not be associated with the concept of a weapon in semantic memory; such a connection may instead be contructed in working memory. Finally, an exemplar may not easily be identified as belong to a category, but the association may be made in the right context. A refrigerator may not be seen as a weapon, but after watching a Jackie Chan movie where a refrigerator door is used in a highly unusual fashion, this connection could be made. A typical RIF study considers the first kind of pairing to indicate a strong semantic association, the second kind a weak semantic association, and the last to be no association. A key point this paper raises is that these different associations may lead to different kinds of processing. The strength of a semantic association may influence the degree to which learning is semantic, and the degree to which it is episodic.

Memory - Cues

Subjects who are aware that they must try to recall words learned at study can use any cue that is best associated with the to-be-remembered words at test. A subject does not have to use the cue they are given. An ideal cue in a generate-recognize framework is one that is available, (the person has no problem remembering the cue itself) is strongly associated with to-be-remembered items, and is unique (it isn't associated with too many items). Normally, with category-exemplar pairs, the category is an excellent cue for

remembering the exemplars. It is available, it is strongly associated with correct responses, and it limits the number of potential responses (to category exemplars). In the case of Anderson et al. (2000), when a new cue is introduced at testing, it makes sense to use cues given at study to generate items instead. Study cues are more strongly associated with items on the test since they have just been practiced, and it is easy to perform recognition decisions on the basis of test cues for generated words. On the other hand, when independent cues are involved in study as well as test, as in Perfect et al. (2004), and there is a unique cue for each exemplar, it makes much more sense to use the non-category cue at testing when it is provided. Here, there is a strong association between the unrelated cue and the exemplar because of the study phase, and, more importantly, the unrelated cue is only associated with one exemplar, whereas category cues are associated strongly with a number of exemplars. However, in the same study, when category cues are given at test instead of the unique cues, the category cues are more likely to be used due to availability. The unique cues would hypothetically be better, but since they are not presented at test, a subject would have to first recall each individual cue, and then the word associated with it. (A good measure of determining what cue is best is finding the average number of generations the model takes to get the correct answer using each potential cue. The model, using this measure, can predict which cue a person will typically use, assuming that people perform recall optimally). A common assumption of RIF studies is that the cues provided at testing are the ones used by subjects; however, since these cues are not always optimal (in any kind of a generate-recognize framework, at least) this assumption should at least be tested.

Memory – Orthography and Phonology

In current RIF studies, the practice phase involves a verbal response, and knowledge of word form is emphasized. Subjects typically need to give responses that are constrained by the first 2 letters of a particular exemplar, and this must cause them to evaluate the word form of the exemplars they recall. Study of the words does not require a response, and involves the highly automatized process of reading. Testing is most similar to practice, in that it involves a verbal response as well. The fact that both practice and testing involve giving a verbal response leads to a greater likelihood that practice words will be given at test (this is also an example of the *generation effect*, see Slamecka & Graf, 1978. In the generation effect, generating a word leads to a greater chance of it being given as a response.) Also, since practice involves word form, and an association between the word form of the category and the word form of the exemplar is hypothesized to take place in the model, the form of the category should be a good cue, and not just the meaning of the category. (This is an encoding specificity sort of effect, as in Tulving & Thompson, 1973).

Prefrontal Cortex

Michael Anderson believes that the PFC plays a significant part in RIF effects. His item suppression view has been discussed above. Recent work (Anderson, 2003) has emphasized the role of the PFC in this item suppression when memory items are selected. Anderson compares memory retrieval to the performance of habitual actions. Executive control is required to override a prepotent response; in a similar way, Anderson thinks that executive control is required to override a strong associate of a given cue in order to remember a weaker associate. This executive control takes the form of item specific inhibition (and when items overlap in their semantic representations, the inhibition is

specific to some semantic features, but not others; Anderson, Green, McCulloch, 2000). In our model, representations compete locally, and there is no PFC mechanism required to resolve competition. Softmax competition (a simulation of lateral inhibition) results in all items competing with one another, and item selection occurring probabilistically according to item strength. These competing views could easily be tested: individuals with PFC lesions would be expected to perform differently on RIF tasks in Anderson's view, since they would no longer be able to inhibit strong memory items to recall weak ones. In our view, RIF performance should be similar in PFC deficient and normal groups.

A problem with the Anderson view of item specific inhibition is that it doesn't make computational sense to inhibit items during practice. Hypothetically, assume the PFC is able to inhibit item representations in memory. If the PFC does this during practice, then RIF occurs. Practicing exemplars that occur on the list cause other strong exemplars to be inhibited. At test, this inhibition causes poorer recall for the exemplars that were not practiced. However, it makes much more sense to inhibit items after they have been recalled. This kind of mechanism leads to better recall, since memory items are inhibited only after they have been recalled. Initially, practiced items are recalled to a greater degree than unpracticed items, and strong semantic exemplars are recalled much more than weak semantic exemplars. After items have been recalled, however, they could be inhibited, which would cause the unpracticed items, and/or the weak semantic exemplars to be recalled with a higher probability. In this kind of model, the strength of weak list items in comparison to strong list items is not important, since the strong list items become suppressed after they are recalled. The strength of weak list items in

comparison to competing non-list items determines whether the weak list items will be recalled. RIF studies are interesting because subjects show worse memory performance for some exemplars *when they are trying to recall as many category-exemplar pairs as possible.* A model or theory that artificially limits the number of exemplars recalled is less interesting; if subjects could recall more words using a proposed mechanism at a different time, why don't they do so?

Modifying Anderson's theory to focus on this kind of competition may solve the problem of why Anderson's theory of inhibition leads to poorer recall scores than would seem necessary. Perhaps items (both list and non-list) need to be suppressed at practice so that non-list exemplars don't intrude at test. This kind of explanation would have the benefit of showing how inhibition at practice leads to the greater recall of category-exemplar pairs.

We hypothesize that item specific inhibition does not occur, however. The PFC may instead be recruited during the recognize portion of our model. In the model, the recognition stage was simple, and items were evaluated against a contextual threshold to determine whether or not they were study items, and whether or not they had been given as a response previously. Items that did not meet this threshold in either case are not given as a response; the response is stopped. This mechanism of comparison is not simulated in detail and may or may not involve the PFC. Our model is agnostic on this point. This conceptualization of the recognition stage of the model as a stopping mechanism parallels Anderson's second function of the PFC, that of stopping retrieval (Anderson, 2001). For example, Anderson and Green (2001) ran a study where subjects were trained to say a given word in response to another word (e.g. say FROG when

PEACH was presented). However *no think* subjects were told to not respond with the word when a certain cue was present, and to also not even think about that word. When tested on their memory for the response words, subjects who had had to not think about words during testing performed worse on recall. Anderson & Green interpreted this result as an inhibitory mechanism squashing out no think words during testing. This finding could be simulated in the model as a higher response threshold in the recognition stage; such a simulation would be very much in line with the Anderson & Green theory of the inhibitory mechanism as a function of the PFC. So although item representations were not inhibited in the model, the model supports, or at least does not contradict, the view that the PFC can inhibit recall during the response process.

Recognition, Episodic Memory, and Semantics

Although some RIF studies were simulated with the current model, the results of studies like Anderson and Spellman (1995) can not be so easily explained. Splitting memory into episodic and semantic components was not sufficient to simulate this result. However, this kind of result can be simulated in at least two different ways with one or two additions to the model. For an explanation of these methods, and implications to the model, see Appendix D of this chapter.

Episodic RIF Studies

The current model has difficulty simulating episodic RIF studies. For example, Anderson and Bell (2001) ran a RIF study in which semantically unrelated words were associated with sentence frames. For example, sentence frames like "The actor is looking at a" and "The teacher is lifting a" were associated with words like TULIP and VIOLIN. Subjects studied "The actor is looking at a" TULIP, "The actor is looking at a" VIOLIN,

and "The teacher is lifting a" VIOLIN. They then practiced "The actor is looking at a tu___", and when "The teacher is lifting a v_____" was given as a test cue, RIF was found. So practicing a word with a certain sentence frame decreased the chance of recalling a competing word even when it was cued with an entirely separate sentence frame.

The model can simulate these kinds of results if word form activation is carried over into test in the same fashion as semantic activation was in the above model of the Anderson and Spellman (1995) results. However, the model then has difficulty simulating Perfect et al. (2004) kinds of results. The interesting question, and one that no current model can answer, is why independent test cues such as those used in Perfect et al. (2004) do not lead to RIF, while episodic test cues, like those used in Anderson and Bell (2001) do lead to RIF. These sets of cues are very similar, in that they have no prior semantic relation to the memory items they are presented with. One possible resolution to this issue lies in the differences between the two tasks. In Perfect et al., words are associated with other novel words. The Anderson and Bell study, however, might best be described as a sentence completion task rather than a memory study. Subjects may not associate words with specific sentences, but rather associate words as viable sentence completion candidates. Words that are practiced are stronger candidates than unpracticed words. At test, stronger candidates are generated more frequently, and so frequent generation and rejection of strong candidates with an inappropriate test cue cause weaker candidates to be given less often as a response.

This explanation is best tested empirically, rather than in a model. An interesting study would be to have only two sentence frame word pairs. For example, The actor is

looking at a " TULIP, ", and "The teacher is lifting a " VIOLIN (note that this hypothetical study is identical to the Anderson and Bell (2001) example, except with the "The actor is looking at a" VIOLIN pair removed). If RIF were still found in this case, it would lend support to the above explanation of episodic RIF results. Episodic RIF tasks could be modeled by associating words to a sentence completion task, as well as to specific sentence frames, and an extension of the model could simulate both Perfect et al. (2004) and Anderson and Bell (2001). If no RIF were found, it would lend support to the item specific inhibition view, but would leave the question as to why Perfect et al. test cues do not lead to RIF, while Anderson and Bell test cues do. No current model would be able to explain this discrepancy.

Comparison with the Norman model

There is one other simulation of RIF effects that merits discussion (Norman et al., 2007). In the Norman et al. simulation, a wide variety of RIF effects were modeled. Items were represented in a distributed fashion, and representations overlapped. The key to the model was an oscillating inhibitory mechanism, which allowed competing memories to be inhibited, and (parts of) target memories to be strengthened. For a given presentation of an item (like APPLE in the presentation of FRUIT-APPLE), those semantic features which made up the concept of an item were strengthened, and similar (and thus competing) features were inhibited. Since practiced items were presented more often then non-practiced items, the net effect was for practiced items to be given as a response more often, and non-practiced, yet related items to be inhibited. The Norman model differs from Anderson's view in that there was no explicit executive mechanisms involved in competitor weakening, and there was strengthening as well as inhibition, but in other

respects the model was an implementation of Anderson's item suppression theory, especially in regards to item-specific inhibition of strong competitors for a given cue.

Our model differs from the Norman et al. model in a number of ways, but two are very significant. First, the Norman et al. model used a form of item inhibition to produce its RIF effects. This item inhibition idea is prevalent in the literature (see introduction), and so the Norman model described a pre-existing theory in terms of specific, detailed mechanisms. Our model was not based on this item inhibition view, and represented a novel theory of RIF effects. It focused on the differing contributions of contextual and semantic factors (similar to Perfect et al, 2004), as well as lateral inhibitory processes during recall. It acted as an existence proof for a theory that was previously thought to be unable to explain critical results, and described this theory in terms of explicit, neurally plausible mechanisms.

Secondly, the Norman model was a specific model of RIF. RIF results in the Norman model were due to regular oscillations in feedback inhibition. This kind of inhibition has not been used to explain other memory findings. Our model was a general memory model that was designed to simulate a number of free recall findings, as well as semantic strategy use. The mechanisms within the model were the same as those hypothesized to produce a wide range of memory findings (e.g. Gilbert & Becker, in preparation), and were based on pillars of the memory literature such as episodic/semantic distinctions. The model was originally designed to simulate free recall and semantic strategies. In order to simulate RIF effects, no additions were made to either the structure of the model or mechanisms within it (in fact, the model was simplified, since, for example, no PFC layer was required to simulate RIF results). The assumptions

on which the model was based were taken from the general memory literature, and the model works using simple, biologically plausible mechanisms. If a specific memory phenomenon can be explained by well established memory processes, why is a specific inhibitory mechanism necessary?

A Category Cue Theory of Semantic Strategies – Chapter 4

When a free recall task is performed, recall is almost always organized in some fashion, as demonstrated by subjective organization studies (e.g. Tulving, 1962). In such studies, even if the presentation order of memory items is varied across multiple study trials of a free recall experiment, when participants attempt to recall items, they do so increasingly in the same order. One widely studied organizational strategy is semantic clustering. If the study list contains words with some underlying semantic structure, then this semantic structure may be used to organize the words during recall (Bousfield, 1953). For example, if a list of words is drawn from a small set of categories (like fruits and vehicles), words of the same category tend to be recalled consecutively (e.g. all the fruit words, then all the vehicles). This semantic clustering in free recall may be due either to the implicit cueing effects of inter-item associations or to an explicit encoding and/or retrieval strategy - the detection of some rule, mnemonic code, or relationship that may serve to categorize/organize the list (Allen, Puff, & Weist, 1968).

The prefrontal cortex plays a crucial role in this organizational ability. Studies of individuals with lesions to prefrontal areas have implicated the PFC in the organization of free recall of long lists, where there is an opportunity for strategic processing, while prefrontal lesions have minimal effect on tests of cued recall or recognition (e.g. Stuss et al 1994; Kopelman & Stanhope, 1998). People with PFC lesions exhibit a significant decrease in the amount of organization in free recall (as measured by subjective organization), and more specifically, a decrease in the amount of semantic clustering. (Hildebrandt, Brandt, & Sachsenheimer, 1998; but see Alexander, Stuss, & Fansebien, 2003 for an alternative view, and the conclusions section of this paper for possible

explanations for this discrepancy). Also, in contrast to healthy controls, such individuals show no statistically significant benefit for recalling lists of related words in comparison to lists of unrelated words (Hirst & Volpe, 1988).

Moscovitch (1994) explains these kinds of findings by assigning frontal (PFC) areas a flexible, executive function. Under this view, the PFC supplies cues to a memory module residing in the Medial Temporal Lobe (MTL). (A related view has been put forward by Frith (e.g. Fletcher, Shallice, Frith, Frackowiak, & Dolan, 1998; Nathaniel-James, Frith, 2002). In Frith's view potential memory responses are "sculpted" by the PFC, which has the effect of biasing some memories over others.) According to Moscovitch, organization deficits in free recall for individuals with PFC lesions are due to an inability to use semantic cues. Note that this is a deficit in the spontaneous use of semantic cues; individuals with PFC lesions who are given explicit and detailed instructions on how to use semantic strategies show equivalent performance to healthy controls (Hirst & Volpe, 1988). Although the PFC has been linked to the use of semantic cues, the exact mechanisms by which the frontal areas supply cues to the MTL are not understood.

One theory of semantic strategy use, then, is that the PFC supplies a category cue to the MTL, which has the effect of organizing recall in terms of semantic clusters. Becker & Lim (2003) proposed a theory of semantic clustering in free recall based on modeling work. The PFC module in this model used a semantic organizational mnemonic without any explicit training; the model "learned" to use semantic organization in the course of simulated free recall tests. However, human subjects almost certainly make use of existing semantic knowledge when using a semantic strategy and do not learn to

perform the strategy itself when performing a free recall task. Also, the semantic cues used by the model were implicit, whereas people may be able to use semantic category labels or semantic associations between items explicitly to cue recall. So although this model provides an interesting explanation of semantic strategy learning, it may not generalize to typical free recall tasks involving words belonging to pre-existing categories. In order to describe how people perform strategies using pre-existing semantic relationships, a new theory and model is required.

Although the PFC is a complex and multifaceted brain region whose numerous functions would be difficult to capture within a single model, a simplified neural network model may be able to capture some of the key aspects of PFC function critical in controlled memory use. While there is not widespread agreement on what exactly the contribution of PFC may be to memory, a key idea put forward by Miller & Cohen (2001) is that the PFC works by biasing the responses of various brain areas. For example, a dominant response in a particular situation may be overridden by the PFC so that a correct, weaker response is performed instead. In free recall of a word list with an underlying semantic structure, those cues that a person would use by default might be overridden by the PFC so that semantic cues could be used instead. This theory of PFC function has the advantage of being easily incorporated within a mathematical or connectionist memory model.

For example, Cohen & Servan-Schreiber (1992) proposed a simple model of PFC involvement in the Stroop task, in which word colours and word names were represented in separate units. Links (weights) between input word units and word name units were stronger than those between input word units and word colour units. The model thus

responded with the word name instead of the word colour when presented with a word. In order to respond with the word colour when presented with a Stroop word (e.g. the word green printed in red) a control unit was required to bias activation in the colour pathway so that it became stronger than the word pathway. This control unit, labeled colour, was active when the task was to name the colour of the word, and was not active when the task was to name the word itself. This simple model captures the three important properties of the PFC: 1) Biasing of other brain areas 2) Active maintenance of this biasing function across the entire length of the task 3) Updating of PFC representations such that when the task ends, the biasing function ends.

The model developed here was an application of these modeling principles applied to semantic strategy use. Instead of a task unit that was used to bias word colour, category labels were kept active in PFC, and were used to bias categorical cuing of memory over other kinds of cuing (like contextual). This categorical cue was kept active, not across the entire task (the recall session, in this case), as was done in the Cohen model, but only as long as it generated successful memory items. Similarly to the Cohen model, this PFC representation was no longer kept active when it was no longer useful. (Note that this kind of top down biasing is similar to that in the model of Becker & Lim (2003) as well. The main difference between the current model and the Becker/Lim model is that the Becker/Lim model learns categorical representations on the fly, and uses implicit activation in the PFC to bias recall, whereas the current model generates categories from existing semantic knowledge, and these categories are hypothesized to be explicitly kept in mind.)

In the following work, we test a theory of semantic strategy use, and examine a model of memory based on this view of PFC function in semantic strategies.

There are a number of functions of PFC in controlled memory use that have been hypothesized (for a full discussion, see the thesis introduction): 1) It has been suggested that the PFC supplies semantic cues for strategic memory retrieval in free recall tasks involving lists of semantically related words (e.g. Shimamura, A.P. 2002). The PFC is also hypothesized to supply other kinds of cues as well. 2) An important executive function in memory may be to inhibit memory representations (e.g. Anderson, 2003) 3) The PFC may also be involved in post-retrieval processing in memory tasks (e.g. Moscovitch, 2002). The following model investigates only the first property of PFC function, that of supplying semantic cues during free recall tasks, for lists of semantically related words. We hypothesize that the PFC does not have to supply cues in an all-or-nothing fashion, rather, that the PFC can provide retrieval cues so as to bias some representations or brain areas more than others, potentially using a number of different cues simultaneously.

Before hypothesizing what PFC mechanisms underlie the deployment of semantic strategies, it is necessary to delineate what the PFC is doing, exactly, and when it does it. When do people notice an inherent semantic structure in a memory list in a free recall task? When do people learn the cues that the PFC will use during free recall? What exactly are the cues that the PFC uses in a semantic strategy? How are these cues generated during recall? Are there a number of ways people can perform semantic strategies, or is there only one way they are typically performed? How are semantic strategies best performed? Does semantic strategy use always lead to semantic

clustering? Are semantic strategies performed in conjunction with other strategies? The following work attempts to answer such questions.

Strategy Use, Semantics, and Clustering

In studies of semantic strategy use, a high degree categorical clustering in an individual's recalled words is often taken as evidence of semantic strategy use, and the absence of categorical clustering is thought to mean that no semantic strategy is used. For example, in Hildebrandt, Brandt, & Sachsenheimer (1998), those individuals with PFC damage showed reduced clustering in comparison to normals; the conclusion is that healthy controls demonstrate greater strategy use as evidenced by their higher clustering scores. Self report data, however, tell a different story. When people are asked about what strategy they use, the use of semantic strategies is not always linked with semantic clustering.

For example, in several of our unpublished pilot studies (unpublished, and listed as Study A, Study B, and Study C below), we collected detailed accounts from participants of the sort of strategies they used in performing a CVLT-like free recall task (Delis, Kramer, Kaplan, & Ober, 2000), which employs standardized lists of 16 words drawn from 4 categories, in 5 blocks of alternating study and recall trials. These accounts revealed that participants adopt a wide range of semantic strategies that would not necessarily result in high semantic clustering scores. In pilot study C, out of 10 participants in the control group, two used a semantic strategy that lead to semantic clustering, and both reported using category labels to organize their recall. "Using category labels" involves generating the name of a category in memory, and using this name as a cue to generate list items. The rest of the participants did not demonstrate

semantic clustering in their recall, but their self-reports indicate the use of semantic information in at least four cases. One participant reported to have used category labels to generate items, but decided to recall all words in serial order as well. Two participants noticed the categorical structure but after the first recall trial concentrated on words missed on prior recall trials. Other participants noticed the categorical structure of the list, but only used categories as a cue when having difficulty remembering words, or used previously recalled words (inter-item associations) instead of category labels, or used completely non-categorical strategies. Finally, some participants failed to notice the categorical structure of the list during recall, and so did not use a clustering strategy and did not evidence any clustering

In the above example, only 2 out of the 6 strategies that involved the use of semantics lead to high semantic clustering scores. Both of these strategies involved using explicit categorical labels to cue memory at recall. These strategies were also "pure" semantic strategies in the sense that they were not combined with other recall strategies. In general, only "pure" semantic strategies where categories are used exclusively as recall cues lead to high clustering scores. Take, for example, the following table (Table 4.1).

The above table represents data from 5 different CVLT-like studies conducted in our lab. Studies labeled by a letter are pilot studies. Studies labeled by number will be presented in full later on in this paper. All of the groups presented in the table included participants who performed CVLT-like memory tasks. Participants were presented word lists involving 4 categories of 4 words each, except for Study A, which consisted of 6 categories of 4 words each. Clustering scores for Study A should thus not be compared to clustering scores in any of the other studies without taking this property into account.

Lists were all presented in a word order where words belonging to the same category were not presented consecutively. The second column represents the number of participants in each group who claimed to have used only the semantic relatedness of list words as a basis for organizing recall – "pure" semantic strategy users. (Subject's recollection of their use of strategies was determined in these studies by asking what strategies they used, when they started using them, and then questioning them regarding the semantic structure they noticed in the lists.) In many cases the description of their strategy seemed to indicate that they used category labels as a cue during recall, although in a few cases the self-report data were unclear. Participants were considered to have used category labels to organize their recall if they mentioned categories in their description of their strategy use, and they didn't mention any additional strategies. The third column indicates the clustering scores for those members of the group who claimed to have used a categorical organizational strategy. The fourth column indicates the clustering scores from the members of the group who did not claim to have used categories (only) to organize their free recall. The fifth column shows the p values for two tests of significance. The first p-value represents whether or not the clusterers and non-clusterers had significantly different clustering scores on trial 5, while the second p-value in brackets represents the score for a repeated measures F test examining whether there was a main effect of reported clustering strategy on clustering scores across all recall trials. Overall, this table thus shows the linkage between a subject's self report of a semantic strategy using category cues at recall, and clustering scores. Those participants that choose to use only category cues at recall are those participants that show significant categorical clustering in their recall scores.

Thus, semantic clustering on the CVLT is hypothesized to be due to a particular semantic strategy – one where a subject is aware of a list's semantic structure, and uses the categories that list words belong to as category cues during recall. Also, when semantic clustering occurs, no other strategies besides categorical cuing are used. The above table does a good job in supporting part of this hypothesis. It links a subject's self-report of the use of a category cuing strategy with clustering scores. This indicates that semantic clustering probably did not occur on CVLT-like tests due to automatic associations between list words in a way that a subject was not aware of (words that are extremely strongly associated with one another were never used in the above studies, though, so this kind of organization may still be possible in some cases). However, only subject's self-report data indicated that semantic categories were used as cues during recall, instead of knowingly using, for example, direct semantic associations between list words. Unfortunately, in some cases self-reports were not entirely clear, as some participants only reported that they "used the fact that the words belonged to the same category" to help organize free recall, for example. A more direct indication that participants use category labels to organize their free recall is necessary to support the hypothesis that only a category label strategy leads to semantic clustering on CVLT-like free recall tasks.

Study 1

To provide converging evidence that on CVLT-like tests, pre-existing category labels are used as an organizing principle in free recall, a study was run. If pre-existing category labels are important, then it should be the degree to which words in a list are associated with such a label that is significant, and not the degree to which words in a list

are related to one another semantically. On many lists consisting of words that have a semantic relation with one another, the effects of these two separate associations are not separated out. However, if there were two CLVT-like lists of words that contained the same amount of semantic relatedness between words belonging to the same category, but different degrees of association with a category label, then the category label hypothesis could be tested. Only a word list consisting of words that can easily be given a category label are hypothesized to lead to semantic clustering.

Participants

24 participants completed the study for course credit. All participants were 1st year undergraduates. Participants were required to have no hearing impairments, and to have learned English as their first language.

Materials

Two lists of words were constructed, an easy-to-label list and a hard-to-label list. Groups of words were assigned to one of the two lists from a master list of 40 words. The words on the master list were grouped according to category and there were 10 categories of 4 words each.

All categories of words had roughly the same semantic relatedness according to LSA measures (an average category cosθ of 0.352 for the hard-to-label words and 0.357 for the easy-to-label words.) LSA (Landauer & Dumais, 1997) is a method that can be used to determine semantic similarity in the following manner. First, a word's LSA vector is constructed according to the frequency with which it appears in different contexts (paragraphs of text). Words that appear in the same contexts have similar vectors. The degree to which two words are semantically related is thus defined in terms

of these vectors (specifically, the cosine of the angle (θ) between their LSA vectors). Lower numbers indicate less semantic relatedness, higher numbers indicate more semantic relatedness, with a cosθ value of zero indicating no relatedness, and a cosθ value of 1 being the maximum amount of relatedness possible. Note, however, that this measure may encompass more than what is traditionally thought of as semantic relatedness. Also note that the LSA matrices undergo a dimensionality reduction using singular value decomposition.

Average category cosθ calculations were calculated using a function on the webpage http://lsa.colorado.edu/cgi-bin/LSA-matrix.html. In the matrix comparison, each of the four words in a given category was compared to the other three words in the same category. This yielded six word pair comparisons. For each pair of words a cosθ value between 1 and –1 was returned. The average of these six cosθ values yielded a category cos.θ. Each of the two lists, both high and low labelability, consisted of four categories. The average category cosθ is an average of the cosθ values for the four categories that make up the list. No category had a category cosθ that differed from this overall value by more that 0.1.

All list words also had roughly the same lemmatized frequency (in a range from 500 –2500). Word frequencies were calculated using a lemmatized frequency list for the British National Corpus (Kilgarriff, n.d.). This frequency list contained the number of times a word had appeared in the British National Corpus. Words were defined in terms of a dictionary entry. So, for example, "kick", "kicks", and "kicked" were judged to be the same word for frequency counts, but the verb "root" and the noun "root" were judged to be different words.

Groups of words were assigned to either the easy-to-label list or hard-to-label list according to labelability norms. To determine these labelability norms, a group of 10 participants were each given the words from the master list grouped according to category. Beside each group of 4 words, the participants were asked to write in the category name, and to rate how difficult it was for them to come up with this category label on a scale from 0-8, with 8 being very difficult to label. The four groups of words with the lowest scores were assigned to the easy-to-label list (average labelability ranking 1.31, with high consistency in the labels assigned by participants). The four groups of words with the highest scores were assigned to the hard-to-label list (average labelability ranking 2.83, with poor consistency in the category labels assigned to the word groups). The other two categories of words were not used in the study.

Both word lists were constructed so that no two words belonging to the same category were grouped consecutively.

Procedure

Participants were tested individually, and were randomly assigned to either the hard-to-label condition or the easy-to-label condition. The appropriate word list was read aloud to participants at a rate of about one word every 1.2 seconds. After presentation of the list, participants were told to repeat as many of the words as they could remember. Participants were given as much time as they wanted to recall items verbally. However, if no words were recalled after 15 seconds, participants were asked if there were any more words they could recall. If a negative response was given, the next trial would commence. In total, there were five trials of list reading + recall.

After the five recall trials, participants were asked some additional questions. They were asked whether or not they used any strategy to help them in their recall. If participants were conscious of using a strategy, they were asked to describe it, and to tell when they first started using it.

Results

An ANOVA with the between subjects variable of labelability and the within subjects variable of trials was run on number of correct words recalled. There were no statistically significant effects for recall except for the expected main effect for trials, $F(4, 88) = 83.41$, $p < .001$.

Semantic clustering scores were also calculated. Raw clustering was calculated by tabulating the number of times words appeared consecutively that belonged to the same category. The raw clustering scores were then adjusted using the formula from Stricker, Brown, Wixted, Baldo, & Delis (2002):

$$C_{sem} = OC_{sem} - EC_{sem}$$

where C_{sem} = the adjusted clustering score, OC_{sem} = the raw semantic clustering scores, and EC_{sem} = the semantic clustering expected by chance on a given trial. The total number of words recalled often influences the maximum number of words that can be clustered, and adjusting the clustering scores attempts to control for this recall effect. High negative adjusted clustering scores indicate organization due to some property other than semantics, while high positive scores indicate significant semantic clustering.

The expected semantic clustering was calculated using a second formula:

$$EC_{sem} = \left[\frac{(r-1)(m-1)}{N_L - 1} \right]$$

where r = the number of correct words recalled on that trial, m is the number of members of each semantic category on the original list (assuming category size is equal for all categories on the list), and N_L = the total number of list words on the original list.

Examining the (adjusted) semantic clustering scores (Figure 4.1), the easy-to-label participants seemed to demonstrate an increase in clustering in later trials as compared to hard-to-label participants. The results of the clustering ANOVA re-enforces this observation: there was a significant trials X labelability interaction $F(4, 88) = 2.88$, $p = 0.027$. There was also a main effect for trials, $F(4, 88) = 4.7$, $p < .01$.

In the easy-to-label condition, five out of twelve participants reported using a semantic strategy that involved using category labels to organize free recall. When the easy-to-label condition was divided into two groups, one which reported using strategies, one which did not, the strategy reporters had an average Trial 5 adjusted clustering score of 5.7, in comparison to a score of 1.4 for those who did not report using a strategy (see Table 1, Study 1 row). In the hard-to-label condition, none of the participants reported using a semantic strategy.

Discussion

Participants in study 1 who studied easy-to-label word lists demonstrated significantly more semantic clustering in comparison to participants who studied hard-to-label word lists. This effect was due entirely to those participants who reported using a categorical clustering strategy. The hypothesis that the use of categorical labels leads to semantic clustering was therefore supported.

However, although for these particular word lists no semantic strategies employing item-item associations were employed, the use of these kinds of strategies can not be ruled out in general. The categories used in Study 1 had moderate LSA cos.θ scores of around 0.35. Words that are more strongly associated with one another may lend themselves to a semantic strategy where item-item associations are employed. Highly semantically related words might lead to increased semantic clustering for non-strategic reasons as well, as participants might tend to give related words consecutively as responses due to semantic priming.

The easy-to-label condition had only moderate labelability scores; participants scored the easy-to-label words about a 3 out of 8 in terms of their labelability.

Study 1 has shown that category labels alone can account for semantic clustering in free recall, yet this result may not generalize to all CVLT-like tests, never mind all potential kinds of free recall tests. It would be difficult, lengthy, and time consuming to exhaustively test all the different CVLT-like tests that would be relevant to semantic strategies and semantic clustering. A better way of examining strategy use in CVLT-like tests is using a connectionist model of free recall. Such a model could quickly simulate a large number of different CVLT-like tests, and make predictions concerning a wide range of findings. Particularly interesting and novel predictions could then be tested in an empirical study. In aggregate, simulation results of free recall could form the basis for an overall theory of semantic strategy use.

Simulation 1

Basic Model

The base model was composed of four layers: a memory item layer, a context layer, a semantic layer, and a PFC layer. The memory layer stored representations of the to-be-remembered word items. Each unit in this layer represented one word; for example, the word DOG would be represented by a particular unit having an activation of 1 and the rest of the units having an activation of 0. This unit would only be active for DOG, thus the model used a localist representation of words. The context layer was split into two parts: constant context, and temporal context. The temporal context layer contained a distributed pattern of activation representing a temporally changing context, similar to TCM (Howard & Kahana, 2001). At any point in time, this layer had units with activity levels between 1 and 0. All of these units taken together (the activity vector) represented the current context of the model. The constant context layer represented stable aspects of the experiment (i.e. whether or not the subject was studying words, or being tested). The semantic layer contained localist representations of semantic category labels. The PFC layer contained units representing strategy use, as well as a rudimentary working memory. All layers were implemented as rate-coded neural circuits and there was full interconnectivity between layers. (see Figure 4.2)

The model can be divided into two main parts. There is a TCM part that consists of the context layer and memory layer, and a second part consisting of semantics and PFC layers that are added to this "base" model. The "base" model is described in Gilbert & Becker (in preparation), an implementation of TCM (Howard & Kahana. 2001). This model was used for two reasons. First of all, it is based on empirical results that control for strategy use. Secondly, it handles non strategic aspects of free recall such as recency better than existing models (see Howard & Kahana, 1999 for an overview of different

models of recency, and for examples of the empirical results on which TCM is based). Thus, if an individual is not using a strategy during free recall, or if a lesioned PFC is simulated, then this part of the model will more accurately simulate free recall in the absence of strategies than other memory models. The additional semantic and PFC layers allow semantic strategies to be simulated.

Training

Prior to training, every word unit in the memory layer corresponding to a memory list word was associated with a context vector in the context layer (the weights were set so that a value of 1 in a unit in the memory layer exactly retrieves this context vector). The pre-list context vectors were randomly generated vectors that were mutually orthogonal to one another (they had a unit length of 1 and they were all orthogonal to one another). These pre-list context associations were meant to represent an average of the temporal contexts of all the times a particular word has been presented to the model. This simulates a human's prior experience with list words prior to testing. Additionally, non-list word items were also given pre-training contextual associations. The contextual patterns that were trained were mutually orthogonal to one another and the contextual patterns of the list items. For a few of the intrusion items, these contextual patterns were then modified with a very small amount of random noise, causing them to be slightly non-orthogonal. Before training commenced, the temporal part of the context layer was set to a random vector. The constant context was also set to a random vector, representing the study context.

Every word unit in the memory layer was also associated with a category unit prior to study. Category units represented the immediate, super ordinate category a

memory item belonged to. So a word unit representing APPLE would be associated with a category unit representing FRUIT. The weights between category labels and exemplars were set between .7 and .9. Memory units were also associated with other units in the category layer. These units represented other semantic associations of a given memory item. For example, APPLE would be associated with PLANT, RED, and a number of other semantic associates other than FRUIT. These other semantic associates of a word unit were not simulated comprehensively. A given word unit had four or five other associates. Weights between these units were set between .1 and .9.

Training the model simulated subject's learning during the spoken presentation of a list of 16 words during a free recall task. Word 'items' were presented to the model by setting the word unit in the memory layer that corresponded to the word item to 1. All other units in this layer are set to 0. This active unit was associated with the current context (the current pattern of activation in the context layer) by using a Hebbian learning rule on the connection weights between the two layers.

$$\mathbf{W}_t = l(\mathbf{c}\mathbf{i}^T) + \mathbf{W}_{t-1} \quad (1)$$

(Where w is the weights between the context and item layers, l is the learning rate, **c** is the current context, t is the current time step and **i** is the item layer activation.)

At the same time as this Hebbian learning took place, the temporal context layer updated its context representation. This new context was a function of the old context activation and a retrieved context

$$\mathbf{c}_t = f\mathbf{r}_i + \sqrt{(1-f^2)}\mathbf{c}_{(t-1)} \quad (2)$$

(where f is a context change parameter and $\mathbf{r}_i$ is the retrieved context for item i.) The retrieved context was calculated by multiplying the active word unit and the weights between this unit and the temporal context layer.

$$\mathbf{r}_i = \mathbf{W}\mathbf{i} \tag{3}$$

(where **i** is the activation in the item layer and **W** is the weight matrix between the item and context layers) Note that the constant part of the context layer did not undergo this updating by retrieved context. Thus, temporal context changed due to the presentation of list items to the model, but constant context contained the same pattern of activation (representing study) throughout the presentation of all list words.

At the same time as the retrieved context was calculated, activity in the semantic layer was determined. Activity in the word unit corresponding to the presented list item caused associated semantic units to be active according to the following equation

$$\mathbf{s}_a = a\mathbf{w}_{as} + \mathbf{s}_{t-1} \tag{4}$$

(where **s** is the activation in the semantic layer, $\mathbf{s}_{t-1}$ is the previous level of activation in the semantic layer, a is the activation of unit a in the memory layer, and $\mathbf{w}_{as}$ is the weights between the semantic layer, and the unit in the memory layer).

Activity in the semantic layer decayed over time. After presentation of a memory item, the level of activity in the semantic layer decreased according to the following equation:

$$\mathbf{s}_a = d\mathbf{s}_a \tag{5}$$

(where d is the decay rate, a number between 1 and 0).

If a particular unit in the semantic layer had an activation level above a given threshold x, then the categorical nature of the word list was said to be detected. This would set a pattern of activity in the PFC representing a categorical strategy.

$$\text{if } \mathbf{s}_u >= x,\ c = 1 \tag{6}$$

Recall

Recall began with a new random pattern of activity representing recall being instantiated in the constant part of context. The course of free recall depended crucially on whether or not the presence of categories had been detected during a previous study or recall session. If no categories were detected, then simulated recall began by having the activation in the temporal context layer cue the memory layer.

$$\mathbf{i} = \mathbf{W}^T\mathbf{c} \tag{7}$$

This cuing resulted in the word units in this layer being activated by a certain amount and the model "choosing" amongst these activated units using the softmax function (Brindle, 1990). The softmax function selected one of the active units in the memory layer and set its activity level to 1 and at the same time it set the other units activity levels to 0. This selection process was done probabilistically, so that units with a greater level of activity were more likely to be chosen then less active units.

$$p_i = \frac{e^{i\mu}}{\sum_{j=1}^{n} e^{j\mu}} \tag{8}$$

(where p is the probability of item a being selected, j is an item in the item layer, a i is the activation of a particular unit, n is the number of items in the item layer, and μ is the softmax parameter). This equation simulated a form of lateral inhibition in which the

inhibition is divisive (shunting) and the unit with the least inhibition has the highest probability of being activated.

The generated word's retrieved context was calculated according to Equation 3. (If a generated word had been presented during study, this retrieved context would include a pattern of activation in the temporal context part, as well as the activation representing study in the constant context.) This retrieved context was compared to the current context (the pattern of activation in the context layer that was used to cue the item layer.) The generated word was said to have been recalled if

$$z_1 < |\mathbf{Wi} - \mathbf{c}| < z_2 \tag{9}$$

(where z_1 and z_2 were upper and lower thresholds for rejection, **c** was the activation in the context layer, and **Wi** was the retrieved context). This equation allowed the model to reject candidate items if they were too similar (i.e., a repetition of the item just recalled) or too dissimilar (intrusion errors) to the cue used to generate them, so that immediate repetitions were prevented, as was selection of words judged to be too dissimilar to the cue used to generate them.

If a word unit in the memory layer was recalled successfully, the temporal context layer was updated using Equation 2. This new context then cued the memory layer, and recall proceeded. Recall ended after a certain number of rejected generations. The model was run a number of times equal to the number of participants it was desired to simulate.

If categories were detected by the model, a categorical strategy was implemented. A word would be generated using Equations 7 and 8. Generating a word also had the effect of activating its semantic representation

$$\mathbf{s} = \mathbf{W}_s\mathbf{i} \tag{10}$$

Semantic representations were activated regardless of whether or not categories were detected; however, this activation only influenced recall when categories had been detected. When a pattern of activity representing a categorical strategy was active, a category unit was selected according to Equation 8. If this category unit represented a category label for the word, rather than some other semantic feature, the PFC maintained the activity level in this unit and both semantics and context were used as a cue for the next recall trial.

$$\mathbf{i} = g\mathbf{W}^T\mathbf{c} + h\mathbf{W}_s{}^T\mathbf{s} \tag{11}$$

(where g and h are parameters that determine the degree to which context and semantics cue the word layer. Both g and h were set to .5 for all simulations).

After a threshold number of unsuccessful generations with this semantic cue active, recall does not stop, but rather, the semantic cue is no longer maintained by PFC, and recall occurs using only contextual cues. If another list word is recalled, its category may be used as a cue similarly as above. Recall ends after a threshold number of unsuccessful generations when no semantic category is active as a cue.

In summary: The model performs identically to Gilbert & Becker (in preparation) in the absence of semantic strategy use. During study, presented words activate semantic features, and sometimes semantic features representing categories are active to an extent that allows the model to detect a categorical organization to the simulated word list. The model then attempts a "semantic strategy". When the first item is recalled, the model may use its category label as a cue for subsequent recalls by storing this label in PFC. When a category cue is "exhausted" after a number of unsuccessful generations occur using the cue, then the cue is discarded, and recall proceeds using only contextual cues. Subsequent

recalls may lead to new category cues being generated and then used for future recall attempts.

Simulations

The model simulated the results of Study 1. The low labelability results are given in Figure 4.3. Here, the simulated word list contained words that, although semantically related, did not belong to an easily identified category. The model did not detect categories, and so Figure 4.3 shows the results of the model when no strategies were used. The number of words recalled increased between Trials 1 and 2, but showed very little improvement thereafter. In comparison, human subjects improved their recall consistently over recall trials. Since the model does not attempt a clustering strategy, the model did not show any clustering above chance levels (this obvious result is not shown).

It is not surprising that the model did not accurately simulate the results of participants in the low labelability condition. Although participants in this condition did not engage in a semantic organizing strategy, all participants reported using some form of organizational strategy. The model, on the other hand, did not use any sort of organizing strategy whatsoever. The model thus acted as a good demonstration why free recall in human subjects always demonstrates some sort of organization (Tulving, 1962). Between the first and second recall trial, contextual associations were learned which make list words more likely to be remembered than extra-list words. Subsequent learning had little effect because the list words were already significantly more strongly associated with contextual cues than non-list words. What is needed is some method of favoring one list word, or group of list words over the rest. This is the point of any organizational strategy. At any given point in recall, using an organizational strategy, some list words are more

likely to be recalled than others. This makes it possible for all words in a list to be recalled, because, at any point in recall, strategy use increases the change of a set of list words being recalled instead of all the list words.

It is possible to increase the number of extra-list words simulated, or to increase the strength of contextual associations between context and extra-list words prior to simulation, in an attempt to have the model show recall improvement across all five recall trials. This change resulted in fewer words recalled initially, though. To counteract this poor initial performance, the threshold for stopping recall must be increased dramatically. This increase in threshold leads to the exact same result for subsequent recall trials: an improvement between trials 1 and 2, and little improvement afterwards. It is possible to get a better fit of the results if, prior to the beginning of recall, the context layer is reset to the pattern of activation that was present at the beginning of study. This property caused the model to recall words in serial order, in a fashion similar to Vousden & Brown (2000).

The high labelability results are shown in Figures 4.4 and 4.5. The model accurately simulated the pattern of results shown by human subjects in free recall. The model showed an increase in the number of words recalled across trials, and demonstrated that, once a semantic strategy is adopted, semantic clustering above chance levels occurs.

Discussion

The model was thus able to simulate basic semantic strategy findings. However, the purpose of the model was not to simulate a comprehensive number of results, but rather to act as an explanatory tool and a way of generating hypotheses that can be tested

with later experimental work. Two examples of explanations of semantic strategy phenomena derived from the model are given below.

Why does the model only use category labels to cue words during free recall and not other semantic features? The exact learning process by which a subject, over his/her lifespan, learns to use category labels as cues during free recall, is beyond the scope of this paper. However, it can be demonstrated in a simulation why subjects use category labels instead of other semantic features. Semantic features are often not necessarily associated with other list words, and there is no systematic way of determining which other semantic features will be useful. For example, APPLE, BANANA, and ORANGE are words on a memory list. All three belong to the category FRUIT. Only APPLE is RED. RED is thus a bad cue for recall.

This kind of recall was simulated using the model above. The results of Figure 4.6 were from a simulation where a semantic feature associated with a list word was also strongly associated with one other list word as well as at least ten other extra-list words. The study phase of the model was identical; recall was similar, with the only difference that only those semantic features that did not correspond to a category label were used as semantic cues. In comparison to the results of Figure 4.4, where category labels were used, it can be seen that these semantic features yield lower recall scores. Presumably, this difference in effectiveness is what people use as feedback to learn to use a categorical organization strategy as opposed to other semantic cues.

Similar results occurred if the list words (category exemplars) themselves were used as a cue in some cases. The model can be set up so that exemplars are highly associated with other extra list words, or not. When a given exemplar had many strong

extra-list associations, and/or relatively weak list associations, it was a bad cue. The problem for subjects is that there is no way of determining beforehand the suitability of a given exemplar as a recall cue. An exemplar may have a lot of associations with extra-list words and few associations with list words, or it may be a very good cue for list words. It would be very difficult for subjects, on a timed free recall test to determine how good a cue an exemplar is, and it seemed difficult to use exemplars in a systematic semantic strategy. However, when subjects were unable to recall any more words, using previously recalled words as cues seemed to be an effective strategy that some subjects used in the pilot studies described above.

Why was the model implemented so that category labels were generated during recall from list words? It might seem more intuitive to have the model notice and then learn categories during study and then use these categories immediately during test. The model predicts that this kind of semantic strategy implementation will not be beneficial, however. This issue is discussed below, and the predictions of the model are shown in simulation 2.

Study 2

Study 1 supports the theory that the degree to which words are associated with a category label crucially affects semantic clustering. However, a significant number of individuals do not attempt to use a semantic strategy at all during CVLT-like tests. Why do some subjects not use a semantic strategy when the opportunity arises? In some cases, subjects may not notice the categorical nature of the study list of words. In the case of the Study C control group (see above), this was the case for two out of the ten subjects. However, another two of the ten participants in this control group noticed the categorical

nature of the list, but decided to use a free recall strategy that did not make use of this information at all, and another four participants used at least one other strategy in conjunction with a semantic strategy. It would be difficult to run a controlled study to determine how subjects choose which strategy they use. However, when the choice of memory strategy is taken away from subjects, at least one reason why subjects do not use semantic strategies becomes apparent. (Table 4.2)

The above table (Table 4.2) includes 4 groups from 3 CVLT-like memory studies. These groups are different from those of Table 1; they all involve a multi-trial free recall task where participants were told, prior to the first study trial, of the existence of a categorical structure in the memory list, and the advisability of using this categorical structure to aid recall. Participants who reported using any strategy that did not involve the use of categories to organize recall were considered to have used a non-category strategy. P-values are reported on the table in a way similar to table 1. The first p-value represents whether or not the clusterers and non-clusterers had significantly different clustering scores on trial 5, the second p-value in brackets represents the score for a repeated measures F test examining whether there was a main effect of reported clustering strategy on clustering scores across all recall trials.

Telling participants of the existence of categories and suggesting they use this property during recall leads to several interesting findings. 1. When participants were told to recall words using a semantic strategy, some participants reported that they did not do so. 2. The amount of semantic clustering for those participants who reported using a semantic strategy was significantly larger then the semantic clustering for those participants who did not report using a semantic strategy. 3. Participants who reported

that they did not use a semantic strategy usually reported trying to at some point, but giving up because they found it too "hard" or "mentally taxing". For example, all participants in Study A who abandoned semantic strategies fit this profile. From this self report data, it appeared that although semantic strategies helped some participants, other participants, when advised (or in the case of Study A, told) to use semantic strategies, found that the use of a semantic strategy hindered their free recall.

So, according to subject's self report data, one reason why semantic strategies were not used was because they were too difficult or unhelpful for some participants. Why do some people find semantic strategies helpful, and use them without being prompted, while other people will not use semantic strategies even when told to do so? Some people may try to use something other than category labels to organize free recall. This happened with three participants in study C who were told the category labels. During study, these individuals concentrated on "other" semantic properties of the list (presumably other shared semantic features, but self-report data is unclear), and then kept trying to use this "other" semantic information at test (See the above simulation for a demonstration of the ineffectiveness of this strategy, and why people might be inclined to discard it). Another explanation for the non-use of semantic strategies is that there is some individual difference variable involved, like intelligence, working memory capacity, laziness, etc. For example, people who have a smaller working memory capacity may have more difficulty using a semantic strategy compared to people with a greater working memory capacity. Semantic strategies might also require more effort in comparison to a strategy where people concentrate on unfamiliar words at study, for example, and so only the more task-motivated might use semantic strategies. The optimal

strategy for memorizing CVLT-like lists may actually be non-semantic; since there are often no significant differences in recall scores for semantic clustering strategies in comparison to non-semantic clustering strategies on these CVLT-like tests, the best strategy may be the one that uses the fewest mental resources.

However, connectionist modeling of semantic strategies led to a different hypothesis. In the simulation of semantic strategies (see above), a very important design decision was that of choosing how the model learned the categories it was to use as cues in free recall. The experimental results were not clear on how this occurs in human subjects. At some point during CVLT-like free recall, a subject notices the semantic structure inherent in the list. Hypothetically, if a subject then decides to use a semantic strategy for recall, he/she will have to memorize the category labels so that they can be later used as cues during recall. There are at least two ways that this can be done. In the current model, the amount of category label learning that took place at study was minimal. Simulated participants took advantage of pre-existing semantic knowledge that linked to-be-remembered words with a semantic category with which they were strongly associated. Recall of one of these words gave the subject the category label for "free". Subsequent list words could then be cued by this category label.

An alternative hypothetical method is for the participants to try and memorize category labels during study, along with the list words. Instead of using word-category associations to be able to recollect categories during recall, this kind of learning would be modeled by increasing context-category associations. During recall, the model would first recall category labels from contextual cues, and then use both category and contextual cues to recall list words. In this theory, contextual cues are crucial in coming up with

category labels; in comparison, the method used in simulation 1 relied on word-category associations. Intuitively, trying to learn category labels at the same time as list words are memorized will be more difficult than trying to learn list words alone. If this assumption is true, then those individuals who try to use a semantic strategy and then abandon it may be using a different implementation of a semantic strategy than those who use it to organize their free recall. Both kinds of individuals try to use category labels as cues to recall list words, however bad strategy users try to memorize or use these category labels during study, while good strategy users use pre-existing category-word associations to come up with category labels during recall.

This hypothesis may seem to be at odds with previous empirical findings. For example, Gershberg & Shimamura (1995) tested frontal patients using a free recall task with a list comprised of categories of related words. Subjects with frontal lesions were given explicit instructions on strategy use at test or at study, and both conditions led to improved recall and clustering scores. This led the authors to conclude that encoding and retrieval were both impaired in frontal patients because subjects could not use strategic cues at these times. In the hypothesis above, it is hypothesized that strategic cues are only useful when used during recall. However, Gershberg & Shimamura provided subjects with the category labels for the lists along with their instructions. Subjects did not have to learn labels during study, and so could encode words in light of these associations. In the above studies, participants must at some point learn or retrieve category labels that are used as cues at recall. It is the process of discovering or learning these cues at study that is hypothesized to lead to a poor semantic strategy implementation in the CVLT.

To test the assumption that semantic strategy use may be harmful when categories are memorized or used in some fashion at study, Simulation 2 was run.

Simulation 2

When implementing a model of a category label semantic strategy in free recall, an important decision had to be made. How and when do participants learn the category labels they use to subsequently cue their recall? Simulation 1 worked under the assumption that participants use preexisting associations between word items and semantic features representing category labels to automatically generate category labels when words are recalled. However, it is also possible for participants who detect an underlying categorical structure in a list of words to memorize the categories along with the words. This kind of strategy would also be necessary if participants did not have pre-existing associations for list words with a category label. In this case, categories would have to be created on the fly, and then used during recall. The results of study 1, where these kinds of low-labelability categories of words on a memory list did yield semantic clustering during free recall, suggest that this kind of strategy does not happen, at least during a free recall task where words are presented at a fairly fast rate (about 1 per second). Why is this result the case? These questions were explored by implementing a model of semantic strategy use in free recall where categories were learned during study along with list words.

Simulation and Results

The model used was the same as in Simulation 1, with a few changes. During study, instead of learning taking place between only the memory layer and context layer, learning was divided between memory-context associations and semantic-context

associations. This modeled participants' memorization of category label cues during study. The modified version of Equation 1 was thus:

$$\mathbf{W}_t = \frac{1}{2} l(\mathbf{c}\mathbf{i}^T) + \mathbf{W}_{t-1} \tag{12}$$

The new learning that took place between the context and semantic layers is represented in the equation:

$$\mathbf{W}(s)_t = \frac{1}{2} l(\mathbf{s}\mathbf{i}^T) + \mathbf{W}(s)_{t-1} \tag{13}$$

(with **W**(s) representing weights between the context and semantic layers).
So instead of learning being concentrated in context-word associations, learning is split evenly between context-word associations, and context-semantic associations.

The model was similar to simulation 1 at test, except for one property. At test, if a categorical strategy was being used, instead of attempting to generate a word item, the model first attempted to generate a category label by using this equation to generate a pattern of activation in the semantic layer

$$\mathbf{s} = \mathbf{W}_s \mathbf{i} \tag{14}$$

and then selecting a unit in this layer using the softmax function (as in Equation 8.) Recall then proceeded using Equation 11. So, if a categorical strategy was used, the model first generated a category, and then attempted to use this category label as a cue for word items (as compared to Simulation 1, where the model first generated a word, and then used this word to generate a category cue for subsequent recall attempts). When a threshold number of unsuccessful generations occurred, a new category was selected using Equations 14 and 8. Recall ended after a threshold number of unsuccessful *category* generations.

Study 1 was simulated with this new model, and the best simulation results are shown in Figure 4.7. Simulation 2 did show semantic organization just like Simulation 1, for the same reasons as Simulation 1. However. Simulation 2 demonstrated significantly fewer words recalled than did Simulation 1 for initial trials. Because simulated participants learned item-context associations to a lesser extent in Simulation 2, as compared to Simulation 1, extra-list category associates of a given category label are much more likely to be given as a response by the model, and then rejected. This led to poorer overall recall performance.

Simulation 2 demonstrated a potential cost to using a categorical strategy. Simulated participants who attempted to memorize category label cues during study while at the same time trying to memorize words, recalled fewer list words. This demonstration serves as a possible explanation as to why some participants did not use a categorical semantic strategy when told to do so, and why some participants abandoned this kind of semantic strategy part way through multi-trial free recall. This result also explains why low-labelability categories demonstrated low clustering scores, even when subjects were told to use categories. Since low-labelability words were not strongly associated with a category label (although they presumably share semantic features), these labels must be created and learned during study. This learning diluted the learning between word items and context, and so would decrease the number of words recalled. Also, in a timed study task like the CVLT, dividing attention between learning words and learning categories might have been very "difficult" when words are presented at a rate of about 1 per second. It may be difficult to rehearse both the list word and a category word

before a new word is presented. Only when words have pre-existing associations with category labels can a semantic strategy be profitably used.

Although the results of Simulation 2 led to a hypothesis that a semantic strategy that yields high clustering scores is due to category labels generated at test, these results only suggest such an explanation. Using the results of the two simulations to generate this hypothesis, however, allows this new theory to be tested in an empirical study. Such a study is now described.

Study 2

The use of simulation and an analysis of a number of CVLT-like studies have pointed to three factors that increase the likelihood of semantic clustering in free recall. First, groups of words in a CVLT-like memory task that are strongly associated with a category label lend themselves to semantic strategy use. Secondly, the use of non-semantic strategies greatly decreases semantic clustering. Finally, particular implementations of semantic strategies may lead to poor recall, and thus subsequently lead to the abandonment of the semantic strategy in mid-task and the adoption of a non-semantic strategy. The first factor, labelability, has already been examined in Study 1. The other two factors have not been directly manipulated in an empirical study, although telling people to use semantic strategies greatly increases the chances that people will try to use a semantic strategy, and then abandon it.

One manipulation that may affect subject's use of strategy is that of interference. Alternative non-semantic strategies require "organizational (PFC based) processing" during study. This was evidenced in Stuss et al. (1994), where subjects performed free

recall on word lists designed to lend themselves to a number of organizational strategies. Those subjects who had PFC lesions demonstrated impaired organization of any kind: pair-frequency analysis revealed significantly less subjective organization for PFC subjects in comparison to normals. Sub-optimal use of a category-label semantic strategy might also consist of subjects trying to memorize categories during study. If an interference task is performed during study, then subjects are hypothesized to be limited to either a useful form of the category-label semantic strategy, or alternative strategies that require "organizational processing" only during recall. The interfering task, in this case, refers to one that involves the PFC in some fashion, so that organizational strategies, dependent on the PFC, either can not be performed or are greatly impaired.

Dividing attention by use of an interference task during a memory task sometimes leads to impaired memory performance and sometimes does not (Moscovitch, 2002). The effects of the interference task depend upon when it is performed, the kind of memory task being performed, and which particular functions an interfering task taps into (Moscovitch, 2002). When a memory task and an interference task both tap into prefrontal areas, then performance of the interfering task will result in decreased memory performance, presumably due to the loss of strategic processing. One such study by Moscovitch (1994) was already performed in conjunction with the CVLT. In this study, Moscovitch tested the free recall of word lists using the standard version of the CVLT (Monday and Tuesday shopping lists). Subjects in his study performed an interfering task in addition to the CVLT and they were divided into groups according to when this interfering task took place: during encoding (when the word list was presented), during retrieval (during free recall of the word list), or during both encoding and retrieval.

Moscovitch found that only in the encoding + retrieval interference condition was recall performance decreased, both in terms of number of words recalled, and amount of clustering.

The current study duplicated many of the procedures found in Moscovitch (1994) (much of the description of the procedure of Study 2 is taken verbatim from the Moscovitch paper). There were several important conceptual differences though. Instead of using standard CVLT lists as Moscovitch did, the labelability lists from study 1 were used. These lists control for semantic relatedness as measured by LSA, and lemmatized frequency, and more importantly, control the ease with which words in the list can be assigned a category label. Study 2 thus contained four interference conditions (interference at study, test, both, or neither), and two list conditions (high and low labelability). For the high labelability participants, it is hypothesized that performance of an interfering task during study will increase the number of participants who use a semantic strategy that leads to semantic clustering, and decrease the number of participants who abandon semantic strategies, or who engage in alternative strategies, compared to participants who do not perform interfering tasks whatsoever. In short, in the high labelability condition, interference at study should improve semantic clustering in comparison to no interference at all. In the low labelability condition, it is hypothesized that participants will not report using any semantic clustering strategy, and no significant clustering differences will be found between interference groups.

Method

Participants

40 participants participated in the study for course credit. All participants were 1st year undergraduates. Participants were required to have no hearing impairments, and to have learned English as their first language. Anyone who was proficient at playing a musical instrument was barred from the study since it was feared that facility performing finger movements would invalidate the interference task.

Materials

The two word lists were identical to the ones used in Study 1.

Procedure

Participants were tested individually. Before being presented with the list of items to remember, participants practiced the finger-tapping task. They were required to tap the fingers of the right hand in the sequence index-ring-middle-small as quickly and as accurately as possible until it was determined that they were proficient (about 3-5 min of practice). At this point, they were introduced to the dual-task procedure. While continuing to tap, they attempted to study a list of 16 unrelated concrete words, none of which contained categories on the easy-to-label list. The list was read at a rate of 1 word per second. At the end of the list, the experimenter tapped the table and called out a number between 100 and 200. This served as a cue for the subject to stop tapping and to begin counting backwards by 7s out loud. After 30s, the experimenter tapped the table again as a signal to begin tapping and to recall as many of the words as possible in any order. The participants were encouraged to use the full minute to recall as many items as possible. Because participants had a tendency to slow down or stop tapping while they were attempting to recall, they were monitored and reminded by pointing to tap in sequence and maintain their pace.

After participants completed this practice phase, they were randomly assigned to one of four interference conditions: tapping at input, output, at both, or at neither. Participants tested during the fall term were assigned to the easy-to-label condition; participants tested during the winter term were assigned to the hard-to-label condition. The finger tapping procedure was identical to the one described for the practice list for the interference at both condition. For the interference at input and interference at output conditions, this practice procedure was also used, except finger tapping was omitted at testing and study respectively. The word list was presented five times with recall following each presentation. After the fifth recall trial, participants were given category labels for the words on the recall list, and told to use these labels to organize their recall.

Results

Words recalled: an analysis of variance (ANOVA) was conducted with trials as a within subject factor, and interference and labelability as between subject factors. There was a significant main effect of trials, $F(4,29) = 159.712$, $p < .001$; and interference, $F(1,32) = 6.020$, $p < .002$. There was a significant trials x labelability interaction, $F(4,29) = 3.917$, $p = .012$; $F(3,32) = 103.442$, $p < .001$; as well as a trials x labelability x interference interaction, $F(12,93) = 2.303$, $p = 0.013$. These findings do not pertain to the hypothesis of the study, but a brief summary of the trends is as follows. High labelability subjects showed increased recall on later trials in conditions where there was no interference at recall, low labelability subjects did not. In general, interference decreased the number of words recalled and as trials increased, so did the number of words recalled, and low labelability subjects recalled more words than high labelability subjects on early trials.

Clustering: as with recall, an analysis of variance was run with trials as a within subject factor, and interference and labelability as between subject factors. There were significant main effects of trials, $F(4, 128) = 4.255$, $p = .003$; labelability, $F(1, 32) = 8.209$, $p = .007$; and interference $F(3, 32) = 7.049$, $p = .001$. Significant interactions included labelability x interference, $F(3, 32) = 5.530$, $p = .004$; trials x interference, $F(12, 128) = 2.994$, $p = .001$; trials x interference x labelability, $F(12, 128) = 2.427$, $p = .007$. The clustering results for the first five trials in the high labelability condition are shown in Figure 4.8. The change in clustering scores after participants were told category labels is shown in Figure 4.9. The clustering results for all six trials in the low labelability conditions are shown in Figure 4.10.

Discussion

Unsurprisingly, participants tended to recall more words across trials. The fact that hard-to-label words were recalled better on early trials than easy-to-label words is very interesting. A possible explanation for this pattern of results is as follows: Participants in the easy-to-label conditions sometimes used a semantic strategy. When they did, only in the interference-at-study condition was it used optimally. When participants did not implement an optimal semantic strategy, their recall scores suffered while they were using the strategy to recall words. In the hard-to-label conditions, participants did not attempt a semantic strategy, and presumably the strategies they did use were always appropriate to the experimental conditions. So, across trials, hard-to-label participants and easy-to-label, interference-at-study participants performed equally, while other easy-to-label participants performed worse because they sometimes attempted to use semantic strategies in a non-optimal fashion.

The expected pattern of clustering results was found. It was hypothesized that the word lists tapped upon existing semantic knowledge, and that due to the timed nature of the CVLT task, it would be easier to generate cues at test than to try and learn them at study. Only in the easy-to-label, interference-at-study condition did participants demonstrate clustering, significantly greater than zero, which supports this hypothesis. Participants in the other easy-to-label conditions attempted to use semantic strategies in five cases total (three in the no-interference condition) according to self-report, but abandoned them in four of these cases; this can be compared with four of five participants using semantic strategies in the easy-to-label condition, interference-at-study-condition, and no participants abandoning semantic strategy use. Lower clustering scores were found due to both less semantic strategy use, and semantic strategy use that was presumably non-optimal.

All participants were able to use category cues when provided with them on trial 6, and this finding is very significant. It suggests that the difficulty in using a semantic strategy is in finding the right cue. Semantic cues do not apparently have to be pre-existing category labels (there do not need to be well learned associations between the category label and the category exemplars), since participants performed well using made up, very general category labels like RELIGIOUS TERMS, and POLITICAL JOBS. This finding also suggests that if a memory task was untimed, then semantic strategy and clustering results might be completely different. If there were no time limits, people would be able to invent category labels or other kinds of unique cues for groups of words, and so any word list that had words that were semantically related might demonstrate significant semantic clustering.

Interestingly, participants in the hard-to-label condition demonstrated clustering scores above zero in the no-interference condition. The reason for this finding can be found in their self reports. In this condition, two participants did use category labels to organize their free recall. Although words were selected that did not share a category label according to subject norms, participants in this condition paired up words. So, for example, although it was impossible for participants to come up with a good category label for the group of words PRAY, WORSHIP, MONK, POPE, it was possible to come up with a category label for just PRAY and WORSHIP, and a different label for MONK and POPE. This is what two participants in this condition did consistently for all words in the list.

General Discussion

Overview

The main conclusion to be drawn from this work is that strategy use can be task-specific and individual specific. Previous models and theories of memory that have addressed semantic strategy use have done so in a results-focused way. Any theoretical process or mechanism that explained results like semantic clustering and learning across trials was considered to be viable. This focus on results was due to the fact that there wasn't a significant amount of research done on what subjects were doing to achieve these free recall results, so theories were not empirically constrained in this way. This work examines some of the process of semantic strategy use, and demonstrates that there are significant individual differences in strategic performance. Also, this work strongly suggests that there are task-specific aspects of strategy use in the CVLT that may not generalize to other free recall tests. A particular average semantic clustering score and

average recall score can not be assumed to be the result of all subjects performing the same way, and the results of a particular test can not necessarily be generalized across all free recall situations.

Semantic strategies and PFC

Impaired or absent semantic clustering in subjects with PFC lesions has often been taken as a sign of a lack of ability to use semantic strategies. Not all semantic strategies lead to semantic clustering however. When people use an "impure" semantic strategy, one in conjunction with another strategic process, then semantic clustering may not occur. It may seem improbable that PFC-lesioned individuals, unable to use a semantic strategy, can use a more complex, multi-part strategy, one part involving the use of semantics. But consider the following hypothetical case. A subject with a PFC lesion tries to recall as many words as possible without any sort of strategy whatsoever. Once this subject can no longer remember any more words, he/she then starts to re-recall words he/she has already given as a response. These words are then used to cue other words in isolation, and the cued words are checked to determine whether or not they were on the memory list. For example, a subject may remember the word GOLD, give this word as a response, re-recall it after finishing initial recall, use GOLD alone as a cue, come up with the words SILVER, METAL, RING, JEWELRY, reject METAL, RING, and JEWELRY, and identify SILVER as a list word. This kind of strategy might result in no significant semantic clustering or subjective organization (as measured by pair-frequency analysis), as long as the majority of words recalled were those recalled during the first pass, where there is no organizing principle to recall. Intuitively, it is probably not the case that PFC-lesioned subjects are spontaneously performing free recall in this complicated fashion. A

more interesting question is if compensatory strategies like these could be taught to PFC-lesioned individuals, and if these kinds of strategies might improve their recall ability. If the PFC works to overcome existing biases, this function implies that significant explicit instruction could allow PFC lesioned people to perform this kind of strategy, although it might be difficult for them to not perform the strategy when dealing with unrelated memory lists.

Another interesting question for future research concerns the aspect of strategic processing that is impaired by a PFC lesion. Is it the detection of semantic structure within a list of words, the implementation of a semantic strategy when a semantic structure is detected, or both that leads to the loss of semantic clustering in subjects with PFC lesions?

CVLT and Generalizablility

The CVLT and CVLT-like studies differ from typical real world memory tasks in a number of ways. First, the CVLT involves a small number of memory items to be memorized, and each memory item is comprised of only a small amount of information. Real world memory tasks, involving potentially a larger amount of more detailed information to be remembered might be more difficult. A student studying a textbook for a test, for example, often has more layers of semantic classification, much more information to learn, and information that is not relevant to the test to ignore, plus the information is presented in a different sensory modality than the CVLT. Real world memory tasks are also often easier than the CVLT in one regard: most real world tasks have no time limit, while words are typically presented during the CVLT at a rate of around 1 per second.

During CVLT-like tasks, it is hypothesized that implementing a semantic clustering strategy where category labels are learned at the same time list words are learned often leads to the abandonment of this strategy due to its ineffectiveness. In many real world memory tasks, this wouldn't necessarily be the case. If there is no immediate time limit, such as a fast presentation rate of to-be-remembered items, then there might be little cost to learning category labels or any other organizing principle along with the memory items. Learning category labels in this way might simply take a bit more time. Also, whereas semantic clustering in the CVLT involves pre-existing category labels, this wouldn't necessarily be the case in the real world. It is not claimed here that people are unable to generate their own category labels or their own semantic cues, only that to do so on CVLT-like tasks is unfeasible.

Although the above empirical results do not necessarily generalize to untimed memory tasks, many of the simulation results above do involve these more general cases. This illustrates the power of a model in comparison to empirical work – it is much easier to generate a theory by simulating a large number of results at once by a model than it is to run a large number of studies. The theory of semantic strategy use derived from this model predicts that the more information there is to be remembered, the more useful it is to break it down into semantically cued chunks in terms of both forgetting fewer words and remembering words faster. It also predicts that more layers of semantic structure will be helpful in these cases (so that cues uniquely pick out a small subset of to-be-remembered information). Finally, it predicts that although there is a cost in memory performance in timed memory tasks, in the case of an untimed memory task, subjects are able to create categories on the fly and teach these categories to themselves along with

the to-be-remembered memory items without any significant cost as long as they spend enough time learning context-item associations. So although the model was constructed to simulate CVLT results, when generalized to non-CVLT situations, it predicts a different set of results.

PFC and Semantic Strategy Use

Stuss and his colleagues have found that PFC-lesions do not impair semantic clustering (e.g. Stuss et al., 1994, Alexander et al., 2003). They feel that semantic strategies are actually automatic processes that are not dependent on the PFC. This view, of course, differs from the one presented here. There are two explanations for the discrepant views. First of all, Stuss et al. found no difference in clustering scores for PFC-lesioned and normal individuals. This may be due to normals not extensively using semantic strategies. A statistical test demonstrating that clustering was significantly greater than zero for both groups would have been helpful. Secondly, semantic clustering may be generated in a different fashion for PFC-lesioned individuals. In the memory studies above, one consistent finding was that participants always used some sort of organizational strategy. If a subject did not use a semantic strategy, then he/she would use a different strategy, and this different kind of organization would presumably interfere with semantic associations which might otherwise lead to significant clustering. PFC-lesioned individuals do not have alternative organizational strategies available, and so this semantic influence is not impaired. If this is the case, then the variance of semantic clustering scores should be higher for normals in comparison to PFC-lesioned individuals. Normals would have very high clustering scores when semantic strategies were used, and low clustering scores when alternative strategies were used, whereas PFC-

lesioned individuals might consistently demonstrate moderate clustering due to more automatic semantic processes. More research is required on this issue.

Summary of Findings

The above studies have led to a number of conclusions. First, people employ semantic information in a number of different strategies. Only a "pure" semantic strategy where people report using category labels to organize free recall, leads to semantic clustering. When memory lists are equated for item-item relatedness using an LSA measure, and lemmatized frequency, only word lists containing easy-to-label words yield semantic clustering. Category labels can be memorized a number of ways. When people are prevented from trying to memorize category labels during study by the use of an interference task at study, they show greater clustering, and a higher percent chance of using a category label strategy. Semantic clustering is thus concluded to be the result of an individual knowingly choosing to use category labels to organize free recall, and generating these category labels in a way that does not interfere with the study of list items. Connectionist simulation describes this memory strategy in terms of specific neural areas and neuron-like units, and explicitly illustrates in detail how the strategy is performed. Simulation results also demonstrate the usefulness of a semantic clustering strategy, in comparison to both non-strategic memory processing, and different implementations of other semantic strategies.

Contributions of the Thesis – Chapter 5

A model of free recall was created, containing a PFC module, a semantic memory module, and a non-strategic memory module; the latter was based on a neural network implementation of a TCM-like memory mechanism. The TCM module was built as a model of non-strategic recall, the PFC and semantic layers were added with the purpose of evaluating the feasibility of our postulated mechanisms of PFC functions in memory tasks that involve semantics. This model was used to generate a number of original results, and motivated several novel empirical findings.

TCM Results

A TCM-like mechanism was incorporated into a model of full free recall. The TCM model is not a full model of free recall, as it only simulates recall of the first item. In order to simulate recall of more than just a single item, mechanisms for preventing repetition errors and for eventually terminating recall must be added. We investigated two potential mechanisms: a generate-recognize mechanism and a response-suppression mechanism. The generate-recognize mechanism was shown to be superior to the response-suppression mechanism in preventing repetitions; the generate-recognize version of the model easily simulated the temporal dynamics of free recall, and performed better at stopping recall in human-like ways in a number of simulations. However, for the TCM-like mechanism in the model, lag recency was shown to be a product of prior recency recalls, as well as rejected intrusion items, rather than due to any property of the model itself. Even with the addition of the mechanism for preventing repetition errors, TCM alone can not simulate multi-trial free recall. Some additional mechanism is required in order for learning to increase across repeated recall trials.

RIF Results

RIF was simulated using the TCM model with the addition of a semantic layer, and several minor changes to simulate cued recall. Importantly, neither a PFC function nor item-specific inhibition was incorporated into the model. Competition dependence properties of RIF were simulated using this model, a result claimed to be impossible by Anderson et al. (1994) for lateral-inhibitory networks. The model overcame traditional difficulties in simulating competition dependence by claiming both semantic and episodic learning occur during a RIF memory task, and using both forms of learning as expressed by different learning rates in the respective layers. Cue independence results were discussed, and the model was shown to provide the only explanation that can simulate the absence of RIF effects for unrelated words, but the presence of RIF effects when semantically related cues were used in indirect ways.

It is important to make clear the claims the model makes concerning inhibition. The model does not claim that inhibition does not take place at the neural level. In fact, the softmax activation function employed in our model could be thought of as a "soft competition", thereby implementing an approximate form of lateral inhibition. The model also does not dispute the claim that there is layer specific inhibition, or that the PFC might perform functions such as response inhibition. Rather, the model is used to argue against the necessity of a specific form of inhibition as an explanation for the RIF effect; in particular, Anderson and colleagues argue that enduring, item-specific inhibition prevents a particular concept or word from coming to mind during a RIF task. Simulations of the model described here demonstrate that item-specific inhibition is not an exclusive explanation of RIF effects. Instead, a theory that does not include item-

specific inhibition was shown to fit a number of important RIF results best, including effects not explainable by item-specific inhibitory accounts.

Semantic Strategy Results

Memory experiments by Shimamura and colleagues, and related work, have demonstrated a crucial role for the PFC in supplying cues during free recall at retrieval. Exactly what these cues are, and the mechanism by which they are applied has not been discussed. The model proposed here, with accompanying empirical work, attempted to explicate these mechanisms in the context of semantic strategies. Participants' self-reports and the results of a "labelability" study suggest that only related words that correspond to a pre-existing semantic category lead to semantic clustering at recall. The model demonstrated that this is due to a high overlap of semantic features among category words; in using category labels subjects are sure to be using a good cue for all the related list words. It was also found that in a timed free recall task, memorizing category labels at study leads to the abandonment of a semantic strategy. This was shown to occur because studying category labels decreases the amount of learning for list items. Subjects who self-generated category labels from list items during recall used semantics effectively. When subjects were prevented from using strategies during study by performing a secondary interference task, they showed greater semantic clustering than subjects who did not perform the interfering task. The empirical result of a benefit of interference at study supports the claims of the model. Finally, the model demonstrated the benefit of semantic strategies; simulations that did not utilize "strategic" mechanisms showed almost no learning after trial 2. Semantic strategies were shown to decrease

competition from other list words, making it less likely the model would generate previously recalled items, and less likely that it would subsequently stop recall.

Modeling Discussion

Purposes of Modeling

In the previous chapters, we have demonstrated a number of uses of computational/mathematical modeling. Modeling has been used in a way that led to results that would not have been found by purely empirical methods. This thesis demonstrates three ways in which modeling may be used to increase the understanding of psychological phenomena.

1. Modeling leads to interesting, novel and testable predictions.

In trying to model semantic strategy use, explicit mechanisms of recall needed to be hypothesized. In the memory literature for many other kinds of memory tests, modeling has already accomplished this goal. For example, in the serial recall literature, there are a number of competing serial recall models that offer explicit mechanisms of serial recall (e.g. Burgess & Hitch, 1999; Lewandowsky, 1999; Page & Norris, 1998; Vousden & Brown, 2000; to name just a few). These models, in making explicit the mechanisms of serial recall, have highlighted important results that are then seen to be useful in adjudicating between the different theories the models embody. An extreme example of this highlighting process in the serial order literature is the Ranchburg effect. Discovered in 1902, this phenomenon turned out to be difficult for the serial order models to account for, and this has lead directly to a re-examination of the effect in empirical testing (e.g. Kahana & Jacobs, 2000). Modeling work in this case led directly to new empirical results by highlighting a very old phenomenon.

There are few such models in the semantic strategy literature, and so, potentially interesting results are not highlighted in this way. In the empirical literature, there has been very little work on semantic strategy mechanisms apart from the examination of subjects' self reports of what they did during free recall. Brain imaging work (e.g. Stuss, 1994) has implicated the PFC in semantic strategies, and so the time is ripe for initial attempts at modeling these processes. The modeling work of Chapter 4 demonstrates how a preliminary attempt at modeling semantic strategies can immediately lead to interesting empirical results. General empirical work on the PFC was applied to the domain of free recall and semantic strategies in the form of a model. The model was able to generate a hypothesis of an organizational benefit to free recall while performing an interference task. There would have been little reason to test for this unintuitive hypothesis if no modeling work had been done.

2. Modeling tests current theoretical assumptions

In the RIF literature, item-specific inhibition is widely believed to be the mechanism by which RIF occurs. Anderson in some of his early work demonstrated the superiority of this explanation to other classes of explanation, and it has remained the dominant theory of RIF effects. However, the models upon which this original explanation stand are extremely simple ratio rule models. Recent results by Anderson theoretically and by Norman computationally have attempted to elaborate on this item-specific inhibition explanation. However, alternative explanations did not undergo this process of elaboration, and remained based on the original ratio rule model.

The model of RIF effects in Chapter 3 was not originally developed to simulate RIF processes. It was created for the purpose of examining semantic strategies. It did,

however, contain the mechanisms necessary to implement a lateral inhibitory account of RIF. Although simple lateral inhibitory networks alone can not simulate the full range of key RIF effects, neural networks that use lateral inhibition, and that model brain processes in more explicit detail by dividing learning between semantics, context, and other brain processes can. Few people have attempted to explain RIF in terms of lateral inhibition. The work of Chapter 3 shows that a model containing enough detail and plausibility in terms of its functional mechanisms can provide a more plausible and compelling alternative to the dominant, item-specific inhibition theory.

3. Modeling work can extend current theories and models

TCM was a mathematical model that made the claim of simulating recency results better than other, competing models (Howard & Kahana, 1999). However, TCM was incomplete as a model of free recall, and it was unclear how it could be extended to simulate other free recall data, besides the recency effect. The model described in Chapter 2 involved the incorporation of a TCM-like mechanism into a full model of free recall. The work was important in evaluating the claims of TCM, to determine whether the mechanism that was responsible for benefits of excellent recency simulation was also responsible for difficulties in modeling other free recall phenomena. Important qualifications were found to be required to TCM: that it needed modification to simulate multi-trial free recall, and that it worked better with a generate-recognize mechanism then with a response suppression mechanism to prevent repetition errors.

Philosophy of Modeling

A major component of the philosophy of modeling that this work is based on is to use explicit mechanisms and structures like temporal context, or separate semantic, PFC

and contextual layers when there is empirical work on which to base these structures. Models in these cases can "flesh out" this empirical work. Both TCM and lateral inhibition in RIF were "fleshed out" in just such a fashion. Just as importantly, when there was no empirical reason to choose a particular modeling structure or mechanism, it was kept as simple as possible. This had the advantage of keeping the model theoretically neutral in a developing field, made it easier to interpret the results of the model, and helped to make the results of the model generalizable across a number of different potential implementations. The benefits of this approach are discussed below in the context of modeling the PFC.

Much of the work on the PFC concerns assigning functions to various parts of the PFC, understanding how the different parts work together, and defining broad classes of PFC function in regards to the rest of the brain (see the Introduction). There are a lot of different theories of PFC function, and ways of dividing the PFC into functional units, but little agreement. Researchers, with good reason, are trying to discover what the PFC does before they attempt to explain exactly how it does it. However, there is a clear link between the PFC and the ability to use semantic cues during recall (Shimamura, 2002). The model attempted to be a general exploration of how cues are used and how the PFC is involved in semantic strategies. This general explanation can be applied to any of the more detailed models of PFC function, so long as the model contains within it a working-memory-like device that can store, over time, a cue-like representation. If one particular theory of PFC "wins out" over others, or the PFC is definitively shown to have a role other than response biasing in domains besides that of free recall, the results of the model

will still stand. If any of these models had been implemented, or a new model created, then the generalizability of the results would be in question.

Also, although recent neuropsychological work on the PFC has greatly increased our understanding of how the PFC works, more work needs to be done before a model of PFC can be created that is constrained by empirical results. It is difficult at this time to model specific mechanisms of PFC function when there is little agreement as to what the PFC actually does, how many different functions it performs, and the specificity of the functions it performs.

Potential Future Directions

Why do People use Strategies?

Strategy use is thought to improve memory performance. For example, when subjects were prevented from using strategies by interfering tasks (as in the TCM modeling work in chapter 2), they recalled an average of about 4 words, whereas subjects who used strategies (as in the semantic strategy work studies of chapter 4) recalled on average a bit more than 6 words. These results are not directly comparable, though, since the above conclusion is confounded by the fact that different word lists were used in the studies, and most importantly, by the performance of a distracter task by those subjects who were recalling without using strategies. Poorer recall scores may be the result of having to perform a distracter task, and may have nothing to do with the presence or absence of strategy use. It is difficult to examine the memory benefit of strategies since people almost always use them in free recall (at least according to self report, see chapter 4).

Tulving and Pearlstone (1966) attempted to examine the benefit of semantic strategies. In their study, subjects who were given category labels for words in a list recalled more words than subjects who just attempted free recall. However, it is probable that the free recall subjects were using some sort of strategy themselves. Also, the free recall subjects had to generate their own cues, while category label subjects had cues provided for them. It may be that it is the additional difficulty of generating cues that led free recall subjects to recall fewer words, and not the fact that they weren't using a semantic strategy.

One method of examining the benefits of strategic recall that gets around this problem is to examine the effectiveness of a particular strategy in comparison to general memory performance. For example, a classic study by Ericsson and Polson (1988) examined the memory performance of a waiter (J.C.) in comparison to university students. The waiter had a specialized strategy which he claimed could help him remember 20 customer orders without writing them down. These claims were verified in experimental conditions, where simulated customers made random food orders. J.C. had an error rate of 3% compared to the average error rate of 20% for students. This study demonstrated a memory benefit for strategy use. J.C. used a particular strategy to remember orders much better than the average person. However, it is highly probable that the student subjects were using memory strategies of their own. The study demonstrates the effectiveness of a particular strategy over other strategies, then, but does not show why people nearly always use strategies in free recall.

One benefit of the modeling work is that it can demonstrate performance in conditions that do not exist in the real world. People always use strategies, or at least

some organizing principle in free recall, but models need not. When strategy biases are removed from the memory model, performance decreases due to competition from other to-be-remembered items. For example, in a word list of 12 items, during multi-trial free recall, without semantic strategy use all twelve items are associated with similar contextual cues. During recall, a contextual cue is used that activates all list words to a significant degree and competition occurs between them. On early recall attempts, words will be recalled without difficulty, but as recall progresses, the chance of recalling a previously recalled item will be greater and greater. Thus, more and more generation attempts will be needed to recall subsequent words. Practically speaking, since subjects tend not to attempt recall after ten seconds of futile attempts to remember (see Chapter 2), this means that fewer words will be recalled. Subjects who do not use strategies will see an exponential (or hyperbolic) increase in their response times, and when it takes too long to recall a word, they will stop trying to recall. This result is not entirely new, as it has been modeled before in simple mathematical models attempting to explain the temporal dynamics of free recall (see Wixted & Rohrer, 1994). However, this link between strategy use and temporal dynamics is a new one. Semantic strategy use, according to our model, biases some to-be-remembered items over others at any given time. When generating items, the model has fewer highly active candidates to choose from at any given time, and so has a smaller chance of recalling already generated items. Since recall stops after a certain duration of unsuccessful recall, the net effect is to increase the number of words recalled.

It would be interesting to see if this result generalized to other memory strategies. It would also be interesting to examine the temporal dynamics of strategic recall in

empirical work. For semantic strategies, the model predicts that within a particular category, the number of generations per successful recall should increase exponentially. Also, the time between recalls when category cues are switched should also increase exponentially. When the number of words per category is small, and the number of categories is small (like in the CVLT), then all words will be recalled quickly on average. So the last word recalled in the last category used as a cue would be recalled nearly as fast as the first word recalled from the first category. Compare this to non-strategic recall, where the last word recalled almost always required at least ten generations to recall (in preliminary simulations), and in many cases took much longer.

Strategies and Untimed Tasks

The free recall tasks that subjects performed in Chapter 4 were all timed. Subjects were presented words at a rate of around 1 per second, and they were also prompted during recall if they did not recall words after a certain time. Time pressure was an issue that influenced memory performance. In the real world, semantic strategy use is not always performed under conditions of time pressure. There are thus questions over the generalizability of the results of Chapter 4. Results from CVLT and CVLT-like tests may not generalize to untimed free recall tasks, or tasks where the presentation of list words is at a slower rate.

It was hypothesized that subjects who attempted to memorize category labels explicitly at study suffered a performance cost in recall. Since time was limited, spending time during study memorizing labels took away study time of list items. Subjects who used this implementation of a semantic strategy tended to abandon semantic strategy use and switched to another strategy, presumably because the semantic strategy wasn't

working well. Subjects who didn't study category labels and who relied on pre-existing associations between list items and categories stuck with semantic strategies. If there was no time pressure in free recall, then there might be a benefit to studying category labels, rather than a cost. Subjects would have all the time they needed to study both category labels and list items. It is almost certainly the case, then, that interference tasks during an untimed period of study would not positively influence the use of semantic strategies, but would rather decrease the effectiveness of them in comparison to a hypothetical no interference task condition. It is hypothesized that the results of Chapter 4 concerning useful semantic strategy implementation would not generalize to untimed memory tasks.

Memorizing lists of words that belong to categories is an artificial task as well. However, there are real world uses of semantic strategies where time is of the essence. Take as an example, once again, the Ericsson and Polson (1988) study of the waiter, J.C., with superior memory performance due to strategy use. The experimental task that J.C. performed involved listening to an experimenter read out orders from a sheet of paper associated with a newspaper photograph. Taking orders from simulated customers is a timed task much like those of Chapter 4. Like the subjects of Chapter 4, J.C. did not have to spend time memorizing category labels, or cues that allow him to access his pre-existing categorical knowledge. J.C. used external cues (customer appearance) that automatically cued categorical representations (which include likely meal orders). His strategy use was more complicated than this, in that he used visual representations of temperature settings, and initial letters as mental short forms for salad dressings, but the main point is that he minimized the memory work done during study to a large extent.

This is hypothesized to be a key factor in the effectiveness of a strategy during a timed task.

Strategies and Pre-existing Knowledge

In the model, semantic strategy use depends upon pre-existing categorical associations that exist for all list words. It is not necessarily the case that these categorical associations must always exist for strategy use to be effective. A subject could almost certainly create categories of his/her own from properties of the to-be-remembered items when using a semantic strategy. This doesn't happen during the studies of Chapter 4 because free recall is timed. A subject simply doesn't have enough time to discover and create these categorical relations and at the same time memorize list words. This is another instance where the results of the semantic strategy studies generalize only to timed memory tasks.

Most of the alternative strategies used by subjects during the studies of Chapter 4 also had the property of relating to pre-existing knowledge. One subject associated word items with musical notes, several subjects used versions of the method of loci, where list items were associated with items in a visual scene in memory, and other subjects used highly personalized knowledge systems that they linked with list items. A serial order strategy was used in a couple of cases; this strategy of recalling words in the order that they were presented does not use pre-existing knowledge. However, this "strategy" does not necessarily involve the PFC. In Stuss et al. (1994), patients with PFC lesions show significantly lower subjective organization scores than controls. Since serial order leads to high subjective organization, this result indicates that PFC lesions might impair serial order. In contrast, in Hildebrandt et al. (1998), subjects with PFC lesions show greater

serial ordering in recall, as measured by serial order ratio, in comparison to controls. (Subjects with temporal lobe lesions showed an even greater tendency toward serial order. One explanation for these results is that damage to long term memory systems caused subjects to rely more on rote repetition, thus leading to serial order.) It is unclear whether ordering items on memory tasks in the order they were presented relies on the PFC, and thus fits the definition of strategy used here.

Cognitive triage effects (as in Brainerd et al, 1990) are also evident in the self-report of some subjects in the empirical studies of Chapter 4. Subjects reported trying to concentrate on word items they felt they missed on previous recall trials; this "strategy" leads to a pattern of recall where previously unrecalled words tend to be recalled first, then previously recalled items, then previously unrecalled items once again. This method of organizing study during a free recall task also makes use of pre-existing knowledge, although it may not seem to on the surface. A subject studies words to the extent he/she feels the words are not in memory. Meta-level knowledge of a subject's knowledge system is required to make this judgment, although in practice this may be a "low level" kind of familiarity rather than a "high level" kind of knowledge.

An interesting, related question is what would happen in low labelability conditions in Chapter 4 if subjects were given category labels to use. Presumably, these labels would not be directly associated with word list items (a pre-existing system of knowledge), but would share semantic features. Extremely preliminary results run on two subjects indicate that these labels, if given to subjects, may lead to clustering in free recall. It is hypothesized that the utility of these created categories depend upon the

degree to which they overlap with the semantic features in subject's semantic representations. This hypothesis could be explored in future studies.

Learning, Strategies, and the PFC

One drawback to simplifying the function of the PFC in the model is that the PFC part of the model does not learn. It performs its tasks algorithmically; it does not need to learn to use semantic strategies. Unfortunately, developmental data on the PFC in terms of strategy use is rather scarce. Rehearsal strategies begin to be used at about the age of 7 years old (Gathercole, 1998). When memory tests do not lend themselves to strategy use, age differences in memory performance are reduced, implying that differences in memory performance as people age are due to the development of strategies (Hess & Radke, 1981). Not much else is known about the development of strategies.

There are a number of kinds of learning and development that may take place in the PFC. The development of strategies may depend on increased synaptogenesis, or increased or decreased myelination. On the other hand, the development of strategies may lead to these kinds of physical changes. PFC learning may transfer over from other tasks that involve this brain area, or it may not. Learning other memory strategies may make it easier to use semantic strategies, or it may not. Semantic strategies may be explicitly taught to individuals, or they may not. The problem of learning is compounded by the dependence of semantic strategies on semantic areas. A categorical structure must be in place if category labels are used as cues in free recall.

The problem of modeling PFC learning is that this sort of modeling is unconstrained by any empirical data. Any sort of PFC mechanism that learns to use category labels as cues, and to stop using a category label as a cue when it is not useful

would fit the data. It is unclear what new knowledge would be gained if such a learning mechanism were implemented in a PFC model of semantic strategy use.

What the model does demonstrate about learning, however, is the benefit of semantic strategy use. Semantic strategies, according to the model, bias some memory items over other memory items. When candidate memory items are generated, there is therefore much less of a chance of repetition, and subsequent stopping of recall. There is a difference in memory performance between simulated subjects who use this memory strategy, and simulated subjects who don't use a memory strategy at all. A hypothetical PFC mechanism would probably make use of this difference in memory performance to learn to use strategies. This makes it probable that recall results must be kept track of in some fashion in order for semantic strategies to be learned.

Modeling other Strategies

The model of semantic strategy works by "noticing" a semantic structure in a simulated list of words, generating a (semantic) cue at recall, and biasing memory item generation using this cue. Could other strategies be modeled in roughly the same fashion? Potentially, yes they could. For example, a method of loci strategy would involve associating different memory items with an imagined path through a visual scene in memory. The cue would be generated from this scene, and would bias memory items through associations learned at study (in contrast to pre-existing associations for semantic strategies).

One aspect of strategic recall that is not covered by this modeling work is PFC work that is done at encoding, like chunking, or the manipulation of memory information toward a more easily remembered form. For example, trying to remember the letters

"iooooesmfpf" is difficult, trying to remember the letters rearranged as "smoofiepoof" is, intuitively, considerably easier, since the letters are organized in a pronounceable and potential humorous pseudo word. This sort of manipulation of memory information is possible in working memory on the fly, and is not accounted for in the present model.

Response Selection and the PFC

A final question revolves around PFC involvement in post-retrieval processing of memories. In the models, a recognition phase determined whether the generated memory item was suitable for recall. Is this recognition function performed by the PFC in human subjects? Empirical results indicate that it might be. For example, Moscovitch and Melo (1997) found that some PFC-lesioned subjects performed confabulation errors across a variety of memory domains, and implicated deficient post-retrieval processing as a potential cause. If the PFC is involved in the recognition phase, would this invalidate the theory that the PFC works by biasing brain areas against dominant modes of responding? This would not necessarily be the case, although an answer to this question would have to come through further modeling work. The PFC may prevent repetition by biasing a null response activity level in premotor areas, or by inhibition of these areas. It is unclear what areas of the brain are involved in the detection of repetition, whether it be memory areas, response areas, or prefrontal areas. Further modeling and empirical work exploring these issues would be very interesting.

References

Alexander, M.P., Stuss, D.T., Fansabedian, N. (2003). California Verbal Learning Test: performance by patients with focal frontal and non-frontal lesions. *Brain*, **126**, 1493-1503.

Allen, M., Puff, C.R., & Weist, R. (1968). The effects of associative and coding processes on organization in free recall. *Journal of Verbal Learning and Verbal Behavior*, **7**, 531-538.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory & Language*, **49**, 415-445.

Anderson, M. C., Bjork, R. A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, **20***, 1063-1087.

Anderson, M.C., Green, C., & McCulloch, K.C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor model. *Journal of Experimental Psychology: Learning, Memory and Cognition, **26***, 1141-1159.

Anderson, M.C., & McCulloch, K.C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 608-629.

Anderson, M.C., & Spellman, B.A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, **102***, 68-100.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, **79***, 97-123.

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, **49A**, 5-28.

Baddeley, A.D. (2002). Fractioning the Central Executive. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 246-260). New York: Oxford University Press, Inc.

Baddeley, A., Chincotta, D. & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, **130**, 641-657.

Baddeley, A.D., Della Sala, S., Paoagno, C., & Spinnler, H. (1997). Dual task performance in dysexecutive and non-dysexecutive patients with a frontal lesion. *Neuropsychology*, **11**, 187-194.

Baddeley, A. D., & Hitch, G. (1977) Recency re-examined. In S. Dornic (Ed.), *Attention and Performance VI.* Lawrence Erlbaum Associates, Hillsdale, N.J. 647-667.

Baldo, J.V., Delis, D., Kramer, J., Shimamura, A. P. (2002). Memory performance on the California Verbal Learning Test–II: Findings from patients with focal frontal lesions. *Journal of the International Neuropsychological Society*, **8**, 539-546.

Baldo, J.V., & Shimamura, A.P. (1998). Letter and category fluency in patients with frontal lobe lesions. *Neuropsychology*, **12**, 209-226.

Baldo, J. V., & Shimamura, A. P. (2002). Frontal lobes and memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), The Handbook of Memory Disorders (2nd Second Edition), Wiley & Sons, Inc.: London.

Battig, W.F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, **80**,* (3, Pt 2).

Bauml, K. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin and Review*, **5(3)**, 459-463.

Becker, S., & Lim, J. (2003). A computational model of prefrontal control in free recall: strategic memory use in the California Verbal Learning Task. *Journal of Cognitive Neuroscience*, **15**, 821-832.

Bjork R. A, Whitten W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology.* **6**, 173–189.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, **49**, 229-240.

Braver, T. S., Cohen, J. D., & Barch, D. M. (2002). The Role of Prefrontal Cortex in Normal and Disordered Cognitive Control: A Cognitive Neuroscience Perspective. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 428-447). New York: Oxford University Press, Inc.

Brindle, J. S. (1990). in: D S Touretzky (ed.) Advances in Neural Information Processing Systems. San Mateo, CA: Morgan Kaufmann.

Bukach, C. M., Bub, D. N., Masson, M. E. J., & Lindsay, D. S. (2004). Category specificity in normal episodic learning: Applications to object recognition and category-specific agnosia. *Cognitive Psychology, **48**,* 1-46.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, **106**,* 551-581.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1314-1319.

Camp, G., Pecher, D., & Schmidt, H. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review*, **12(3)**, 490-494.

Carter, K. L. (2004). Invertigating semantic inhibition using a modified independent cue task. PhD thesis, University of Kansas, Lawrence, KS.

Clark SE and Gronlund SD (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review* **3**: 37-60.

Cohen JD, Servan-Schreiber D, McClelland JL (1992). A parallel distributed processing approach to automacity. *American Journal of Psychology* , **105(2)**, 239-269.

Delis, D., Kramer, J., Kaplan, E., Ober, B. (2000). *California Verbal Learning Test-Second Edition.* San Antonio: The Psychological Corporation.

Dennis S and Humphreys MS (2001). A context noise model of episodic word recognition. *Psychological Review* **108**: 452-478.

Diamond, A. (2002). Normal Development of Prefrontal Cortex from Birth to Young Adulthood: Cognitive Functions, Anatomy, and Biochemistry. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 466-503). New York: Oxford University Press, Inc.

Dimitrov, M., Granetz, J., Peterson, M. et al. (1999). Associative learning impairments in patients with frontal lobe damage. *Brain and Cognition*, **41**, 213-230.

Ericsson, K.A., & Polson, P.G. (1988). An experimental analysis of a memory skill for dinner order. *Journal of Experimental Psychology, Learning Memory and Cognition, 14*, 305-316.

Farrell, S. & Lewandowsky, S. (in press). Empirical and theoretical limits on lag-recency in free recall.*Psychonomic Bulletin & Review.*

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language,* **51**, 115-135.

Fletcher,P., Shallice,T., Frith,C., Frackowiak,R., Dolan,R. (1998). The functional roles of prefrontal cortex in episodic memory. *Brain*, **121(7)**, 1249-1256.

Fox, J. & Das, S. K. (2000). *Safe and Sound: Artificial Intelligence in Hazardous Applications*. Menlo Park, CA: AAAI Press.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, **47**, 469-497.

Gathercole, S. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, **39**, 3-27.

Gershberg, F.B. & Shimamura A.P. (1995). Impaired use of organizational strategies in free recall following frontal lobe damage. *Neuropsychologia*, **13**, 1305-1333.

Gillund, G., & Shiffrin. R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Glanzer, M. & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behaviour*, **5**, 351-360.

Goldman-Rakic, P.S. (1987). Circuitry if primate prefrontal cortex and regulation of behavior by representational memory. In: F. Plum (ed.), Handbook of Physiology, The Nervous System, Higher Functions of the Brain, Section I, Vol. V., Part 1, Chapter 9 (pp. 373-417). Bethesda, M.D: American Physiological Society.

Goldman-Rakic, P.S. (1991). Prefrontal cortical dysfunction in schizophrenia: the relevance of working memory. In Carroll, B.J. & Barrett, J.E. (Eds.), Psychopathology and the Brain (pp. 1-23). New York: Raven Press.

Goldman-Rakic, P.S. & Leung, H.C. (2002). Functional Architecture of the Dorsolateral Prefrontal Cortex in Monkeys and Huimans. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 85-95). New York: Oxford University Press, Inc.

Goshen-Gottstein, Y., & Moscovitch, M. (1995). Repetition priming for newly-formed associations is perceptually based: Shallow processing and format specificity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1249-1262.

Hess, T. M. & Radtke, R. C. (1981). Processing and memory factors in children's reading comprehension skill. *Child Development*, **52**, 479-488.

Hildebrandt, H., Brand, A., Sachsenheimer, W. 1998. Profiles of patients with left prefrontal and left temporal lobe lesions after cerebrovascular infarcations on california verbal learning test-like indices. *Journal of Clinical and Experimental Neuropsychology*, **20**: 673-683.

Hintzman D (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* **95**: 528-551.

Hirst, W. & Volpe, B.T. (1988). Memory strategies with brain damage. *Brain & Cognition*, **8**, 379-408.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, **25(4)***, 923-941.

Howard, M. W., & Kahana, M. J. (2001). A distributed representation of temporal context. *Journal of Mathematical Psychology, **46***, 269-299.

Hull, C. L. (1943). *Principles of Behaviour*. New York: Appleton-Century-Crofts.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, **30***, 513-541.

Jacoby, L.L., Toth, J.P., & Yonelinas, A.P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, **122***, 139-154.

Jacoby LL, Yonelinas AP, and Jennings JM (1997). The relation between conscious and unconscious(automatic) influences: A declaration of independence. In Cohen JD and Schooler JW (eds), *Scientific Approaches to Consciousness*, pp. 13-47. Mahwah, NJ: Erlbaum.

Janowsky, J.S., Shimamura, A.P., Kritchevsky, M., & Squire, L.R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavoural Neuroscience*, **103**, 548-560.

Johnson, S.K., & Anderson, M.C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, **15***, 448-453.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, **24**, 103-109.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **28**, 530-540.

Kilgarriff, A. (n.d.). lemma.al. Retrieved Dec 28, 2001, from ftp://ftp.itri.bton.ac.uk/bnc/.

Kjeldergaard, P. M. (1968). Transfer and mediation in verbal learning. In T. R. Dixon and D. L. Horton (eds.), *Verbal behaviour and general behaviour theory* (pp. 67-96). Englewood Cliffs, NJ: Prentice-Hall.

Kopelman, M.D., & Stanhope, N. (1998). Recall and recognition memory in patients with focal frontal, temporal lobe, and diencephalic lesions. *Neuropsychologia*, **36**, 785-796.

Lamming, D. (2005). Personal communication.

Landauer, T. K., & Dumais, S. T., (1997). Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

Levine, B., Black, S.E., Cabeza, R., Sinden, M., McIntosh, A.R., Toth, J.P., Tulving, E., & Stuss, D. T., (1998). Episodic memory and the self in a case of isolated retrograde amnesia. Brain, 121, 1951-1973.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition: In B. H. Ross(Ed.), The psychology of learning and motivation (Vol. 43) (pp. 163-214). San Diego, CA: Academic Press.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval induced forgetting. *Psychological Science*, ***12***, 148-152.

McClelland JL, McNaughton BL, and O'Reilly RC (1995). Why there are complementary learningsystems in the hippocampus and neocortex. Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* **102**: 419-457.

Mensink, G. L. M., & Raajimakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, **95**, 434-454.

Miller E.K, & Cohen J,D, (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, ***24***:167-202.

Moscovitch, M. (1994). Cognitive resources and dual task interference effects at retrieval in normal people: the role of the frontal lobes and medial temporal cortex. *Neuropsychology*, *8*, 524-534.

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia. Journal of Verbal Learning and Verbal Behavior, 15, 447-458.

Moscovitch, M. & Winocur, G. (2002). The Frontal Cortex and Working with Memory. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 188-209). New York: Oxford University Press, Inc.

Murdock BB (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial order information. *Psychological Review* **100**: 183-203.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, **1997**, 839-862.

Murdock, Jr., B.B. & Okada, R. (1970) . Interresponse times in single-trial free recall. *Journal of Experimental Psychology, **86(2)**,* 263-267.

Nathaniel-James,D.A., Frith,C.D. (2002). The role of the dorsolateral prefrontal cortex: evidence from the effects of contextual constraint in a sentence completion task. *NeuroImage*, **16(4)**, 1094-1102.

Norman, K. A., Newman, E, & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychological Review*, **114(4)**, 887-953.

O'Reilly, R. C. (2006). Biologically Based Computational Models of High-Level Cognition. *Science*, **314**, 91-94.

Perfect, T. J., Moulin, C. J. A., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition, **28***, 1111-1119.

Perfect, T, J., Stark, L., Tree, J. J., Moulin, C. J. A., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language, **51***, 399-417.

Petrides, M. (1994). Frontal lobes and working memory: evidence from investigations of the effects of cortical excisions in nonhuman primates. In F. Boller & J. Graffman (Eds.) Handbook of Neuropsychology, Vol 9 (pp. 59-82). Amsterdam: Elsevier.

Petrides, M. (2000a). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *Journal of Neuroscience*, **20**, 7496-7503.

Petrides, M. (2000b). Mapping prefrontal cortical systems for the control of cognition. In: A.W. Toga & J.C. Mazziotta (Eds.), Brain Mapping : The Systems (pp. 159-176). San Diego: Academic Press

Petrides, M. (2000c). Frontal lobes and memory. In F. Boller & J. Grafman (Eds.), Handbook of Neuropsychology, Second Edition, Vol. 2 (pp. 67-84). Amsterdam: Elsevier.

Petrides, M., & Pandya, D. N. (2004). The frontal cortex. In the Human Nervous System, G. Paxinos and J. K. Mai (Eds.), San Diego: Elsevier Academic Press, 2nd Edition, Ch. 25, 950-972.

Postman, L. 1971. Transfer, interference and forgetting. In *Woodworth and Schlosberg's experimental psychology*, 3rd ed. (ed. J.W. Kling and L.A. Riggs), pp. 1019-1132. Holt, Rinehart and Winston, New York, NY.

Postman, L. & Philips, L. W. (1965) 'Short-term temporal changes in free recall'. *Quarterly Journal of Experimental Psychology,* **17**, *132-138*

Raaijamkers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.

Rapesak, S.Z., Reminger, S.L., Glisky, E.L., Kaszniak, A.W., & Comer, J.F. (1999). Neuropsychological mechanism of false facial recognition following frontal lobe damage. Cognitive Neuropsychology, 1, 267-292.

Roediger, H.L., & Schmidt, S.R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 91-105.

Roediger, H. L., Balota, D. A., & Watson, J. M., (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neathe, & A.M. Suprenant (Eds.), The nature of remembering: Essays in honour of Robert G. Crowder (p 95-115), Washington, DC: American Psychological Association.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory and Cognition, 21, 803-814.

Rohrer, D., & Wixted, J.T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22*, 511-524.

Schacter, D.L., Curran, T., Galluccio, L., Milberg, W., & Bates, J. (1996). False recognition and the right frontal lobe: a case study. Neuropsychologia, 34, 793-808.

Shallice, T. (2002). Fractionation of the Supervisory System. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 261-277). New York: Oxford University Press, Inc.

Shallice, T., & Burgess, P.W. (1996). Domains of supervisory control and the temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 351, 1405-1412.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 179-195.

Shiffrin R.M .,and Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, **4**: 145-166.

Shimamura, A.P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology*, **28**, 207-218.

Shimamura, A.P. (2002). Memory Retrieval and Executive Control Processes. In Stuss, D.T. & Knight, R.T. (Eds.) *Principles of Frontal Lobe Function* (pp. 210-220). New York: Oxford University Press, Inc.

Shimamura, A.P., Gershberg, F.B., Jurica, P.J., et al. (1992). Intact implicit memory in patients with frontal lobe lesions. *Neuropsychologia*, **30**, 931-937.

Shivde, G., & Anderson, M.C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. D. Gorfein (Ed), *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*, pp. 175-190. Washington, D.C.: American Psychological Association.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 592-604.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behaviour*, **10**, 400-408.

Smith, A. D. (1973). Input order and output interference in organized recall. *Journal of Experimental Psychology*, **100**, 147-150.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology, **10**,* 438-464.

Sternberg, R.J. & Tulving, E. (1977) The measurment of subjective organization in free recall. *Psychological Bulletin*, **84**(3), 539-556.

Stricker, J. L., Brown, G. G., Wixted, J., Baldo, J. B., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test- Second Edition: Background, rationale, and formulae. *Journal of International Neuropsychological Society*, **8**, 425-435.

Stuss, D. T., Alexander, M. P., Palumbo, C.L., Buckle, L., Sayer, S, and Pogue, J. (1994). Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks. *Neuropsychology*, **8**: 355-373.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, **69**, 344-354.

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in paired-associate learning. *Journal of Verbal Learning and Verbal Behaviour*, **1**, 321-334.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words, *Journal of Verbal Learning and Verbal Behaviour*, *5*, 381-391.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, ***80***, 352-373.

van Maanen, L. & van Rijn, H. (2006). An accumulator model account of sematic interference in memory retrieval. In Proceedings of the Seventh International Conference on Cognitive Modeling (pp. 322-327). Trieste, Italy.

Veling, H., & Van Knippenberg, A. (2004). Remembering can cause inhibition: Retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, ***30***, 315-318.

Vousden, J.I., Brown, G.D.A., & Harley, T.A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, **41**, 101-175.

Williams, C., & Zacks, R. (2001). Is retrieval induced forgetting an inhibitory process? *Journal of Psychology*, ***114***, 329-354.

Wixted, J.T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. Psychonomic Bulletin and Review, 1(1), 89-106.

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Studies of directed forgetting in older adults. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 143-156.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Yevgeniy, B. S., Tully, M., Wingfield, A., Kahana, M. J. (2006). Temporal associations and prior-list intrusions in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 792-804.

Appendix A: Competition-Dependence Results (Anderson & Bjork, 1994)

Study: WEAPON-SWORD WEAPON-PISTOL (standard high-association pairs)

Practice: WEAPON-SW__

Test: WEAPON-?

RIF (impairment in recalling PISTOL)

Study: WEAPON-FOOT WEAPON-NAIL (low-association pairs)

Practice: WEAPON-FO__

Test: WEAPON-?

No RIF (no impairment in recalling NAIL)

Study: WEAPON-SWORD WEAPON-FOOT (mix of high and low association)

Practice: WEAPON-SW__ (note: high association practiced)

Test: WEAPON-?

No RIF (no impairment in recalling FOOT)

Study: WEAPON-FOOT WEAPON-PISTOL (mix of high and low association)

Practice: WEAPON-FO__ (note: low association practiced)

Test: WEAPON-?

RIF (impairment in recalling PISTOL)

Appendix B: Cue Independence Results

(Anderson et al., 2000)

Study: FRUIT-APPLE FRUIT-PEAR (standard high-association pairs)

Practice: FRUIT-PE__

Test: RED-?

RIF (APPLE is given less if FRUIT-PE__ is practiced)

(Anderson and Spellman, 1995)

Study: GREEN- LETTUCE GREEN-EMERALD SOUP-TOMATO SOUP-CHICKEN

Practice: GREEN-LE_____

Test: GREEN-? SOUP-?

RIF (Both EMERALD and TOMATO are given as responses less if GREEN-LE_____ is practiced)

(Pefect et al. 2004)

Study1: ZINC-APPLE STOOL-PEAR (unrelated pairs)

Study2: FRUIT-APPLE FRUIT-PEAR (standard high-association pairs)

Practice: FRUIT-PE__

Test1: FRUIT-?

RIF

Test2: ZINC-?

No RIF

Appendix C: Other RIF Results

(Anderson and Shivde, in preparation)

Study: WEAPON-SWORD WEAPON-PISTOL (standard high-association pairs)

Practice: WEAPON-SWORD (note: full practice)

Test: WEAPON-?

Little to No RIF

(Anderson and Bell, 2001)

Study: The teacher is lifting a – VIOLIN The actor is looking at a – VIOLIN
The actor is looking at a – TULIP

Practice: The actor is looking at a tu___?

Test: The teacher is lifting a v_____?

RIF

Appendix D: Simulating Anderson and Spellman (1995)

In Anderson and Spellman, list items have semantic similarity with other list items associated with other cues. For example, in Experiment 2, subjects studied a list of 6 words associated with GREEN, and 6 other words associated with SOUP. Half of the GREEN words and half of the SOUP words were also VEGETABLES. Practicing a GREEN associate that was a VEGETABLE not only reduced recall for non-practiced GREEN words, but also SOUP words that were VEGETABLEs. In this case, practicing word pairs was said to decrease recall for exemplars that were associated with a different cue during study! (The words that show RIF effects (reduced recall) even with a different cue at test will be subsequently referred to using Anderson's nomenclature: NRpS items, for No Retrieval practice, Similar). Two different methods of simulating these results are proposed.

One method involves carrying over semantic activation from study into test. Normally, when a word was presented to the model, its semantic associations would become active and some learning would occur. After this occurred, the activated semantic pattern was simply "dropped" from the semantic layer. However, this semantic activity can be maintained, and decay either due to some decay parameter, or using subsequent semantic retrievals (or not at all). The net effect of this process at test would be to cause the semantic layer to contain activity representing three kinds of semantic features. Most of the activity in the semantic layer was in units representing the semantic features of category cues. If GREEN-LETTUCE was studied, GREEN would be highly active in semantics. A second kind of semantic activity would be non-category semantic features of exemplars. IF GREEN-LETTUCE was a word pair, these non-category features would

be all the features of LETTUCE that were not captured by GREEN. These kinds of features tend to be exemplar specific, so there would be significantly less of this activity in comparison to GREEN activation(since this activity depends on the number of times LETTUCE was presented, which was much less than the number of times GREEN was presented). The third kind of semantic activity was that of the shared feature or category. In this case if GREEN-LETTUCE and SOUP-TOMATO were word pairs, semantic features representing vegetables would be active to a large degree. Here, the activity would be similar to that found with the actual category cues. Both VEGETABLE features and GREEN features were presented the same number of times, so this makes sense.

The semantic activation that was carried over from study to test can be used to cue memory items, along with the category cue. This does not result in RIF for NRpS items in related conditions. It actually leads to an increase in NRpS recall in unrelated conditions. This result was exactly what was found in Perfect et al.'s (2004) analysis of Anderson & Spellman (1995). When VEGETABLE was used as an implicit cue due to this carry over effect, in related conditions, both LETTUCE and TOMATO were activated due to the carry over, and they competed with one another to be recalled. In unrelated conditions, TOMATO did not compete with LETTUCE, and so showed a greater increase in recall. This modeling work can be surprisingly complicated, and the accompanying explanations become very involved. Preliminary, simple simulations can demonstrate the effect, however.

The model was changed by using not only category cues at test, but also a pattern of activation representing other category features present during study (such as VEGETABLE in the above example.) This pattern of activation represented the semantic

activation carried over from study; the actually process was not simulated. Implicit category features (like the ones which represented VEGETABLE) were three times as strong as other non-category features, since, for a given list category, three of the six items contained a NRpS category, and only around one item on average would be associated with other non-category semantic features. Practiced implicit category exemplars had larger weights associated with the given implicit category features than unpracticed implicit category exemplars, to simulate the extra learning during study. In practice conditions, the activation of implicit category features was increased as well. So at test, list items were cued by the semantic representation of the current category cues, semantic representations of other studied category cues, and implicit category cues. As mentioned above, the model shows an increase in recall for NRpS items in the no competition condition. When there is no competition, the implicit category cue increases the chance of generating a list word that is associated with both the study category cue and the implicit category cue. So a word cued by both VEGETABLE and GREEN is more likely to be recalled than a word cued just by GREEN. When there is competition, practiced implicit category cue items were more likely to be recalled than unpracticed implicit category cue items. If LETTUCE is cued by VEGETABLE and GREEN, and is practiced, and TOMATO is cued by VEGETABLE and SOUP, and is not practiced, then the extra practice LETTUCE undergoes will cause it to be recalled more for two reasons. First, LETTUCE has stronger weights associated with VEGETABLE features due to the practice. Second, because of the practice, VEGETABLE is a stronger implicit cue during study than it would have been if no practice occurred. TOMATO does not see as significant a gain in recall due to softmax competition with LETTUCE.

The model can also simulate Anderson and Spellman (1995) with a second method: by building on its recognition decision making process. Items with contextual and semantic associations similar to those of items already given as a response might be incorrectly rejected as repetitions. With this method, it is not the cue that is used that is important, but rather the degree of semantic and contextual similarity between different test words. For example, TOMATO and LETTUCE are both semantically similar, and so if both were studied items, recalling one of these words may decrease the chance of recalling the other due to increased repetition rejections, regardless of the cue used to generate the item.

There are two problems with this solution, however. First, it is necessary for NRpS items to be more similar to one another than category exemplars for the model to work as it did before. For example, if the similarity of EMERALD and LETTUCE was greater than the similarity of LETTUCE and TOMATO, then any semantic similarity comparison that rejected TOMATO would also reject EMERALD. If NRpS items were not more similar to one another than category exemplars were, then the model would have two sources of RIF effects: lateral inhibition and repetition prevention. With two sources of RIF effects, the model would have to be re-evaluated on all the above simulations. On the other hand, if NRpS items were not similar enough to one another, then they would not be rejected as repetitions, and the model doesn't work. Secondly, when NRpS items were too similar to already recalled exemplars, the model prevented them from being given as a response all of the time, rather than a decreased percentage of the time. So RIF effect sizes were too large.

A closer examination of recognition can lead to a solution to both of these problems. Normally, contextual information was used by the model to prevent repetitions. Semantic similarity was not a good basis for deciding whether or not a generated item appeared on the list, or whether it had already been recalled. Since a category cue was used at test, it was highly likely that such a cue would generate semantically related words. However, since the category cue is given at test in the RIF paradigm, when performing recognition decisions, subjects may be able to discount semantic information that is related to the category, and focus on non-category-cue semantic information. Subjects know that any item they recall should be related to the category used to cue it; the non-category-cue semantic information then becomes crucial in determining whether or not the item is a repetition. Most of the time in RIF studies, non-category-cue semantic features will not conform to any pattern; the chance of several cued items sharing features is small. However, in Anderson and Spellman-like studies, both retrieval practice items and NRpS items share non-category-cue semantic features. Retrieval practice items have both semantic and word form information available to them to make recognition decisions (according to the model) whereas NRpS items have only non-category-cue semantic information. Because NRpS items share some of these features with retrieval practice items and other NRpS items, when a NRpS item is generated by the model, it may be rejected as a repetition.

To understand how such a process would work, divide the semantic layer in a hypothetical version of the model into categorical features and non-categorical features. During recall, a given categorical cue clamps a certain pattern of activity in the semantic layer. This activity represents the current categorical features. This information is not

useful for recognition decisions, since the use of the category as a cue always causes this activity to be present. What is useful is all the non-categorical features active in the semantic layer. These features are not activated by the cue at test, and so any activity must be due to previous recalls, or residual activity from study. The model can either be modified so that semantic activity persists across recalls, or more simply, use the pattern of activity generated by the last recalled item to make the comparison. If there is a high degree of similarity between the current non-categorical semantic activity and the non-categorical semantic activity of a generated item, the item would be judged as a repetition. Again, in this kind of model, it is the non-categorical semantic features which are important to recognition decisions. These are precisely the kinds of features that are activated by the implicit categories in Anderson and Spellman (1995).

A very simple version of this model was run. It was identical in all respects to the normal model, except that it did not use categorical features in its recognition decisions. If the model used the pattern of semantic activity generated by the last recalled item to make repetition error judgments, then the model accurately simulated Anderson and Spellman. However, this kind of model also produced a significant number of repetition errors. Although no repetition error data is available concerning Anderson and Spellman, the number of repetition errors this kind of model produced was almost certainly too many.

Thus, a new version of the model was created to solve the issue of repetition errors. It had semantic activity persist during study and test. This activity did not cue memory items during recall, but rather, was used in recognition comparisons. In order for NRpS items to be rejected some of the time, but not all of the time, an additional

component was added to the model. A retrieved context layer was added, which acted similarly to retrieved context in TCM (Howard and Kahana, 2001). A description of this process lies beyond the scope of this paper, so the function of this layer will only be described briefly. The layer adds noise to the recognition process. List words were associated with pre-existing retrieved contexts randomly; the amount of similarity between the contexts of any two words was random. If the activity in this layer changed according to retrieved context, and this layer was used to make repetition error judgments along with semantics for generated items, then NRpS items were rejected as repetitions some of the time, but not all of the time. It may seem unparsimonious to add an entirely new layer to simulate a single result. However, the retrieved context layer was always a part of the model. It was used to simulate non-strategic free recall results in other work(see Gilbert and Becker, in preparation, for a full description of how the retrieved context layer works). It was not included initially in RIF simulations because it had no effect on results, other than to increase the variance in the results.

Both methods of modeling Anderson and Spellman require some sort of semantic "priming". In the carry over method, semantic features that were activated by list words remain active to a certain extent and cue memory during testing. The recognition method requires that the pattern of semantic activity that a generated item is compared to be influenced by previous recalls.

The very simple versions of these models presented here are meant merely as an existence proof to demonstrate that the model can handle difficult RIF results. RIF research has been dominated by an item-specific inhibition view, and this work attempts to demonstrate the viability and existence of other approaches, but does not claim to be

the final word. This paper focuses on RIF results that can be obtained when episodic and semantic memory processes work independently in recall. Additional work could further explore recognition decision processes, or expand on semantics. Although the current model requires additional components to be able to simulate these results, and the Anderson theory (Anderson, 2003) very parsimoniously accounts for them, the current model remains the only one able to explain why related list words and implicit cues lead to RIF effects in a large number of cases, but the use of unrelated cues (as in Perfect et al., 2004) does not. Also, the simulation of free recall findings require these additional components as well, and so adding these components results in the ability to simulate a wider range of memory findings than Anderson's theory can simulate (Gilbert & Becker, in preparation).

Appendix E: Word List for Semantic Strategy Studies

Word List: Easy to Label	Word List: Hard to Label
carrot	monk
opera	fridge
goat	congressman
cart	rug
bull	mayor
blues	fork
potato	pope
truck	stool
pop	pan
bicycle	sofa
deer	governor
corn	pray
folk	bowl
money	lamp
beans	worship
taxi	treasurer

Appendix F: Model Parameters

The model had a temporal context layer of 80 units, a study layer (constant context) of 10 units, and an item layer of 100 units. There were 72 simulated subjects. The model had a 50 word vocabulary.

Parameters in the model:

Learning rate (l) = 1

Context change parameter (f) = .707

Upper threshold for rejection (z1) = 0.05

Lower threshold for rejection (z2) = 0.25

Number of rejected generations needed to stop recall (n) = 5

Softmax parameter (π) $\approx$ 2

For the response suppression model, a grid search was done using the following parameters: learning rate, context change, softmax, and response threshold (z). Bias (b) was set at -1, and decay (d) was set at .9; these parameters influenced repetition errors, but not total number of words recalled or temporal dynamics.

For the horse rate model, parameter searches were made using the following parameters: learning rate, context change, response threshold(z), time step (ϖ), and noise (η)

No parameters are given for the RIF model, since a general case version of the model is shown which demonstrates key effects to be parameter independent. No parameters are also given for the semantic strategy model, since the model was used to provide explanations and make predictions. These predictions and explanations, once generated, do not require a model to understand them, and they are tested with empirical work.

Table 3.1

Hypothetical Kinds of Learning That Take Place During RIF Studies with High and Low Related Exemplars

	Amount of Learning during the Study Phase	Where Learning is Concentrated during the Practice Phase
High related exemplars	high semantic (but small learning rate due to well learned links), low context	high semantic (but small learning rate due to well learned links), low context high word form
Low related exemplars	high context normal semantic	high context, normal semantic high word form

Table 4.1

Free Recall Clustering Results for Participants who Used Categories as Cues, and for Participants who did not Use Categories as Cues

Study	Number of participants who said they used categories	Average trial 5 adjusted clustering score for clusterers	Average trial 5 adjusted clustering score for non clusterers	Significance of difference between clusterers and non clusterers
Pilot Study A, control group	10/20	9.9	2.4	$p<.001$ ($p<.001$)
Pilot Study B, control group	4/8	8.4	0	$p<.01$ ($p = .001$)
Study 1, easy to label group	5/12	5.7	1.4	$p<.01$ ($p < .001$)
Pilot Study C, control group	2/10	9	-.9	$P < .001$ ($p<.001$)
Study 2, no interference group	1/5	8.2	-0.8	$p = .001$ ($p = .001$)
Total/Average	22/55 (40%)	8.2	0.4	
Study 2, interference at study group	4/6	4	0.2	$p = .059$ ($p=.046$)

Table 4.2

Free Recall Clustering Results in Studies where Participants were Told to Use Categories as Cues During Recall

Study	Number of participants who did not use semantic strategy	Average trial 5 clustering score for clusterers	Average trial 5 clustering score for non clusterers	Significance of difference between clusterers and non clusterers
Study A, experimental group	7/20	7.3	1.1	P < .001 (p = .002)
Study B, experimental group	3/8	7.4	0.8	p = .003 (p = .063)
Study C, told of semantic relatedness group	3/10	6.8	-0.3	p = .011 (p < .001)
Study C, told category labels group	4/10	6.4	0.2	p = .004 (p = .005)

Figure Captions

Figure 2.1

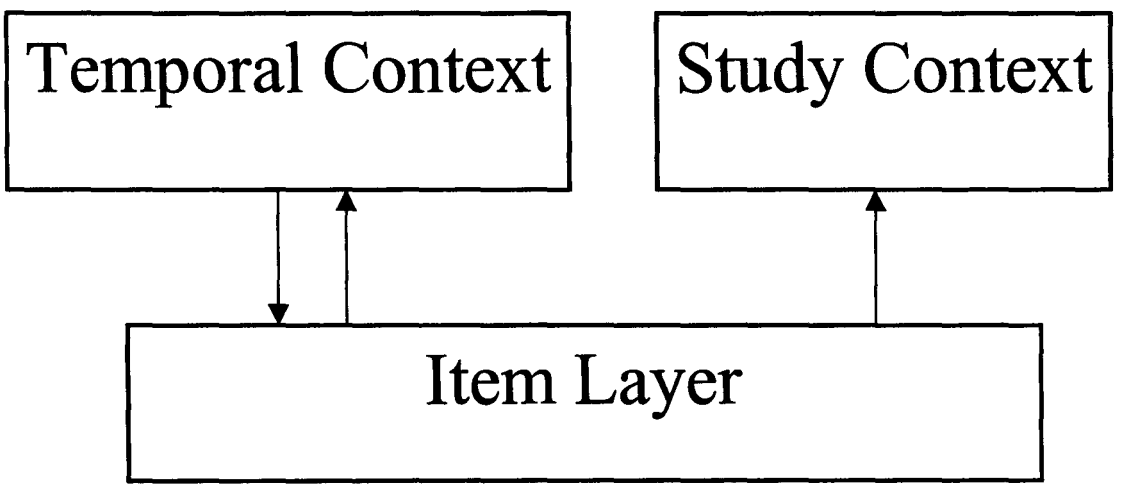


Figure 2.2

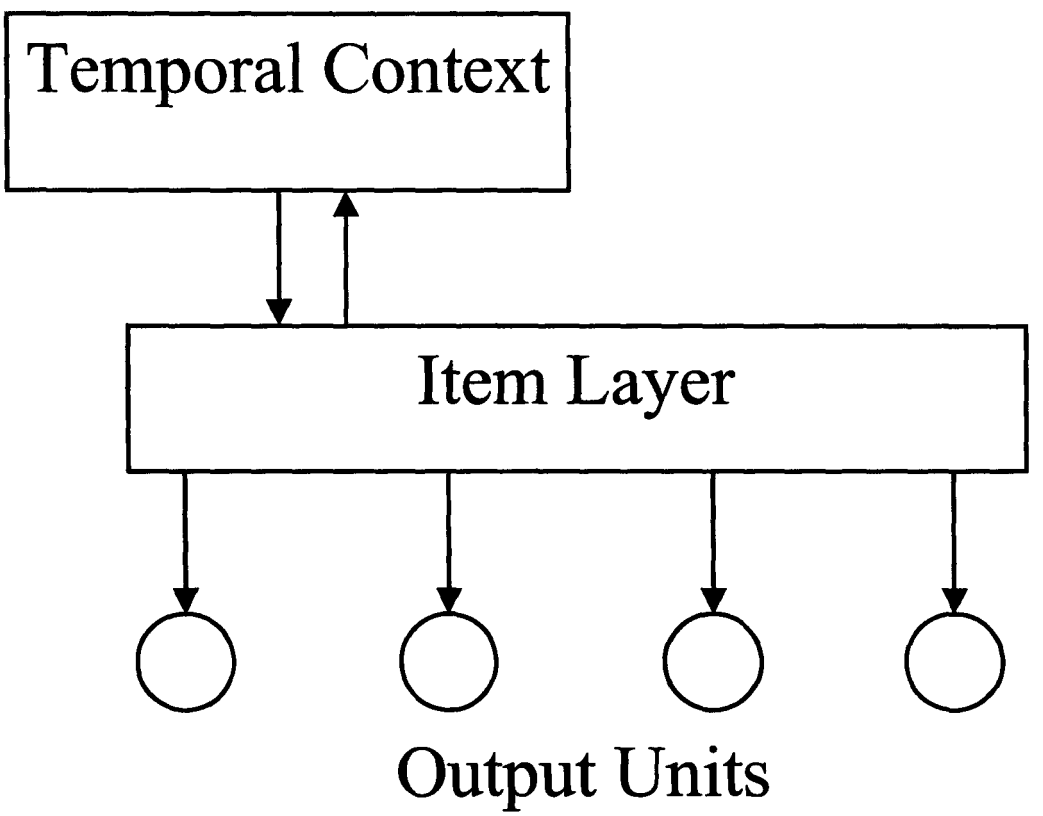


Figure 2.3

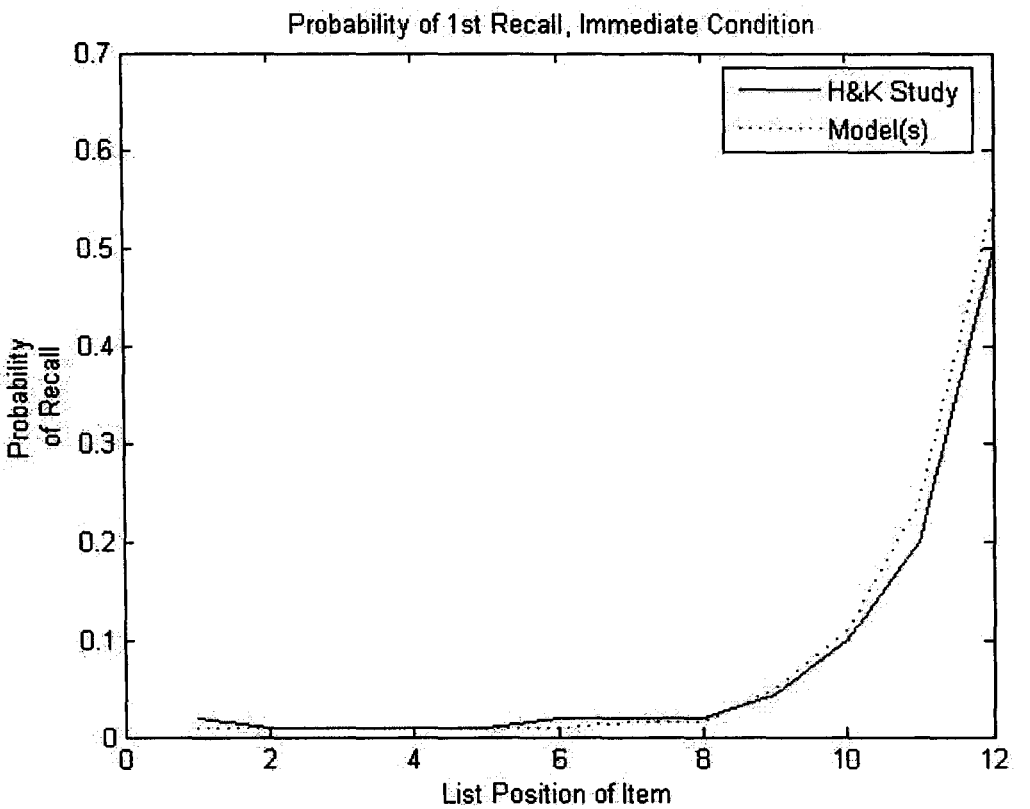

Probability of 1st Recall, Immediate Condition
H&K Study
Model(s)
Probability of Recall
List Position of Item
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
2
4
6
8
10
12

Figure 2.4

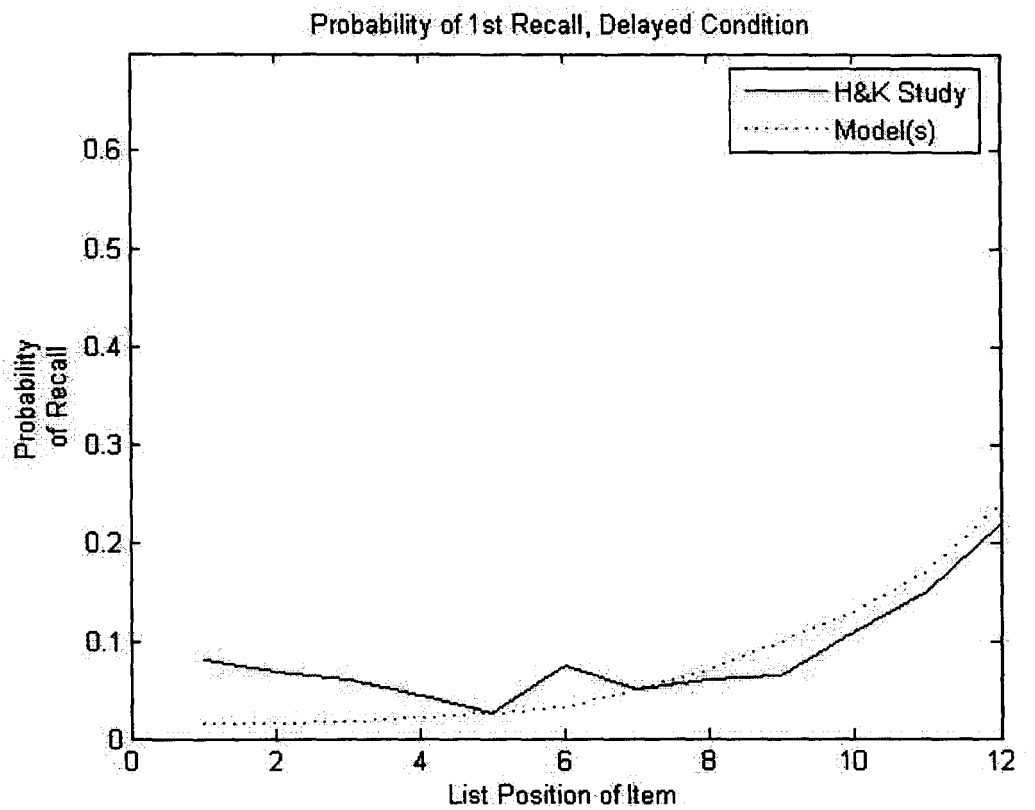

Probability of 1st Recall, Delayed Condition
H&K Study
Model(s)
Probability of Recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0
2
4
6
8
10
12
List Position of Item

Figure 2.5

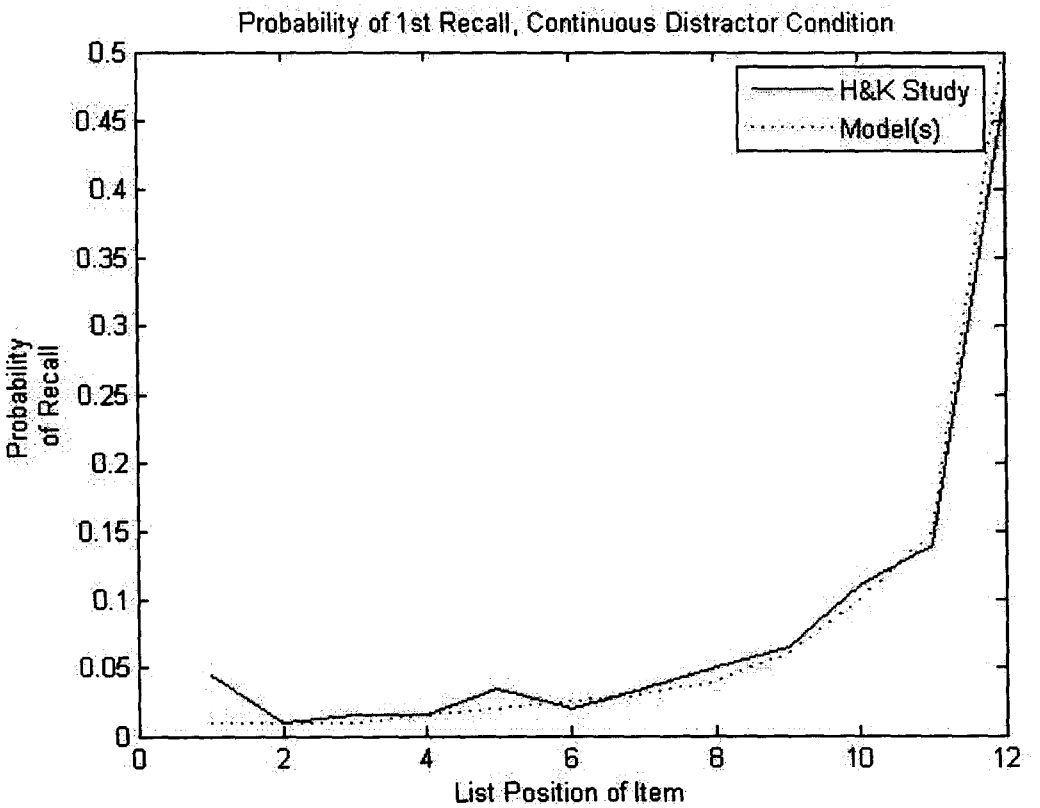

Probability of 1st Recall, Continuous Distractor Condition
H&K Study
Model(s)
Probability of Recall
List Position of Item
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
2
4
6
8
10
12

Figure 2.6

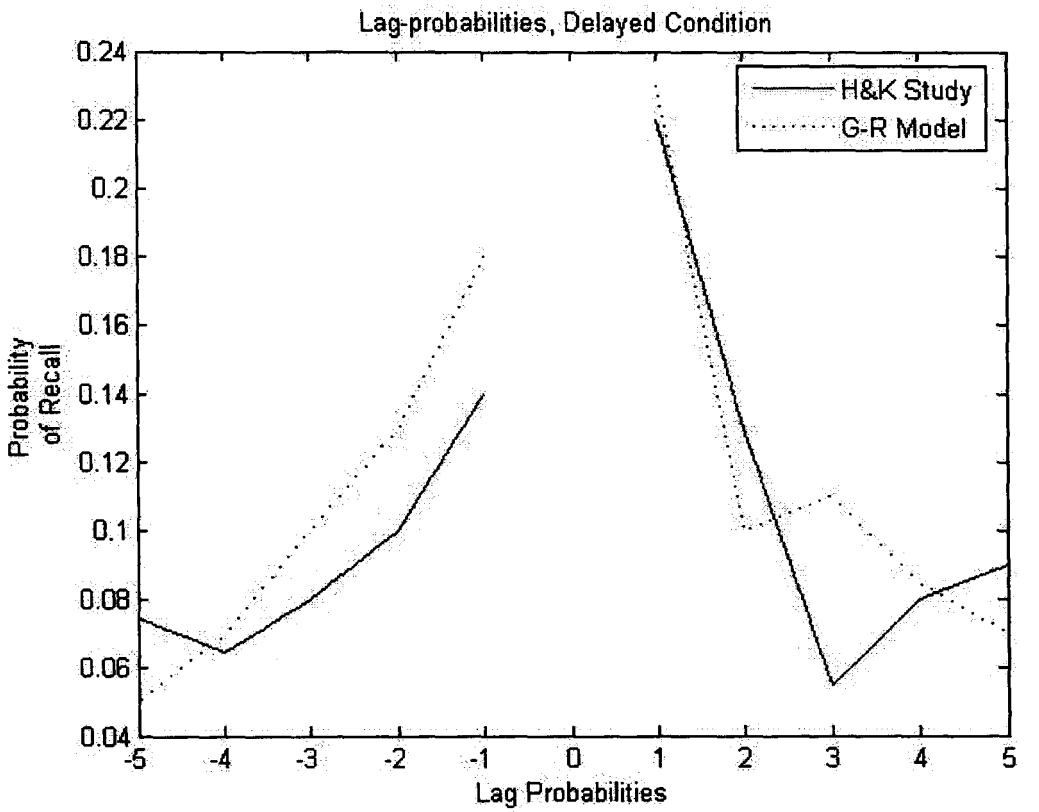

Lag-probabilities, Delayed Condition
H&K Study
G-R Model
Probability of Recall
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
-5
-4
-3
-2
-1
0
1
2
3
4
5
Lag Probabilities

Figure 2.7

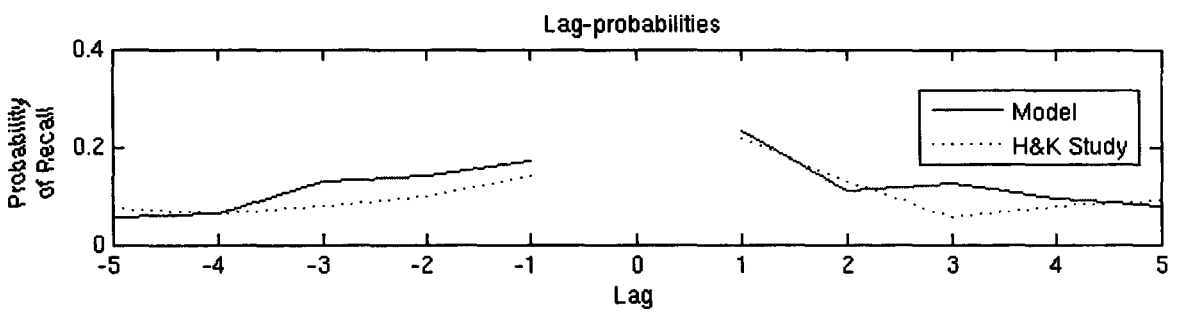

Lag-probabilities
Probability of Recall
0.4
0.2
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
Lag
Model
H&K Study

Figure 2.8

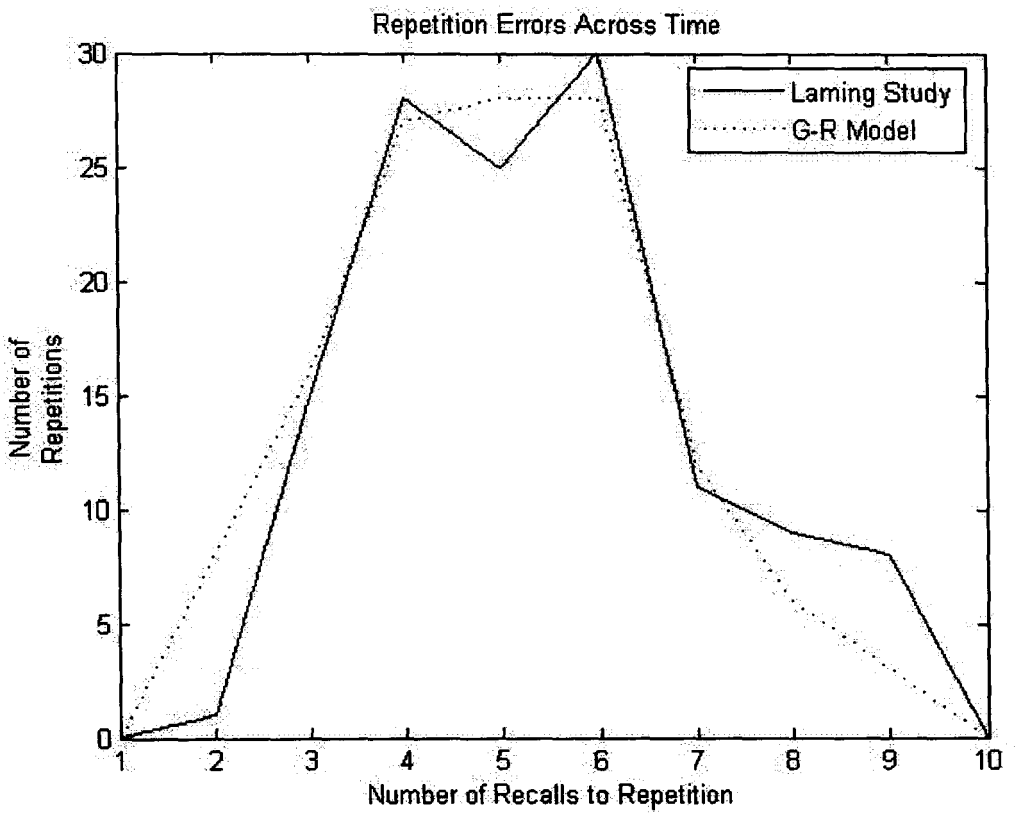

Repetition Errors Across Time
Laming Study
G-R Model
Number of Repetitions
Number of Recalls to Repetition
0
5
10
15
20
25
30
1
2
3
4
5
6
7
8
9
10

Figure 2.9

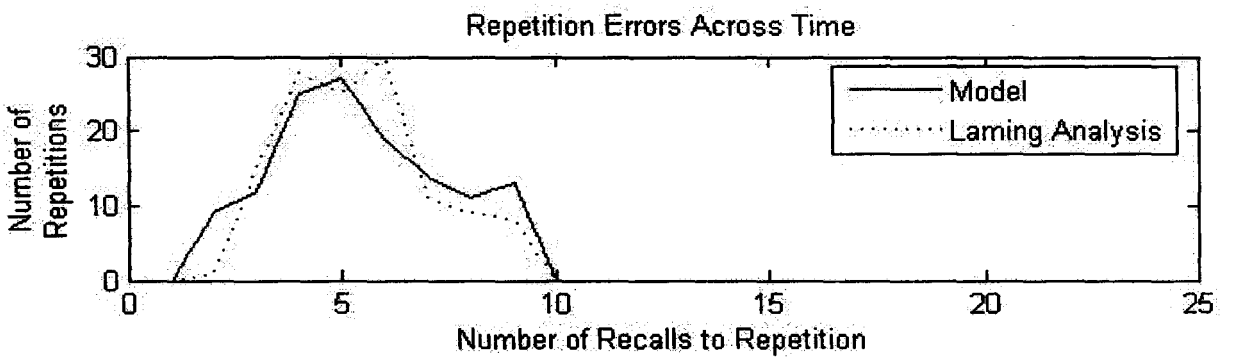

Repetition Errors Across Time
Number of Repetitions
30
20
10
0
0
5
10
15
20
25
Number of Recalls to Repetition
Model
Laming Analysis

Figure 2.10

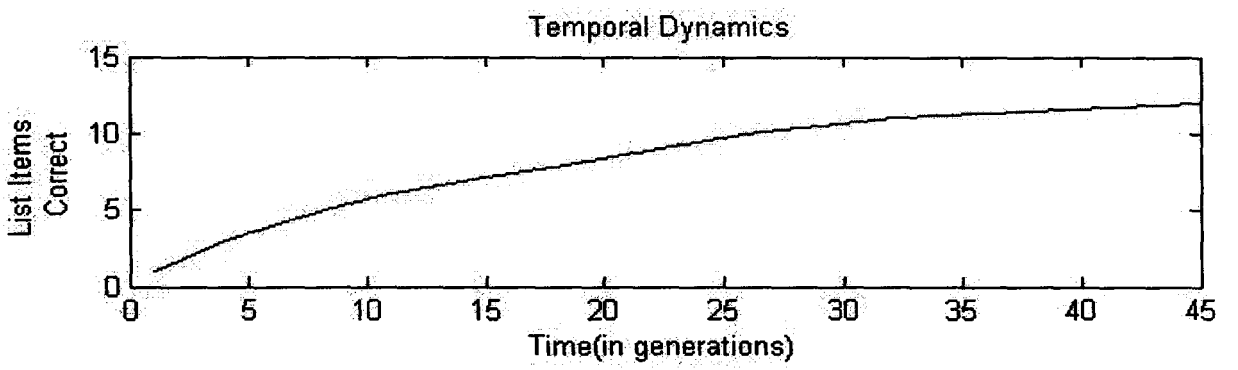

Temporal Dynamics
List Items Correct
15
10
5
0
0
5
10
15
20
25
30
35
40
45
Time(in generations)

Figure 2.11

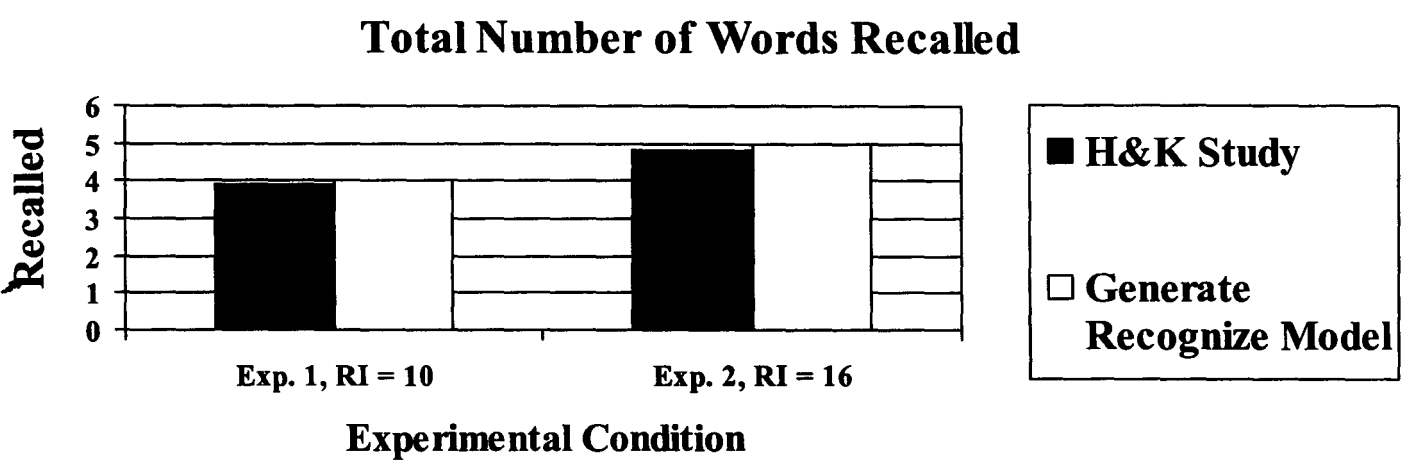

Total Number of Words Recalled
Recalled
6
5
4
3
2
1
0
Exp. 1, RI = 10
Exp. 2, RI = 16
Experimental Condition
H&K Study
Generate Recognize Model

Figure 2.12

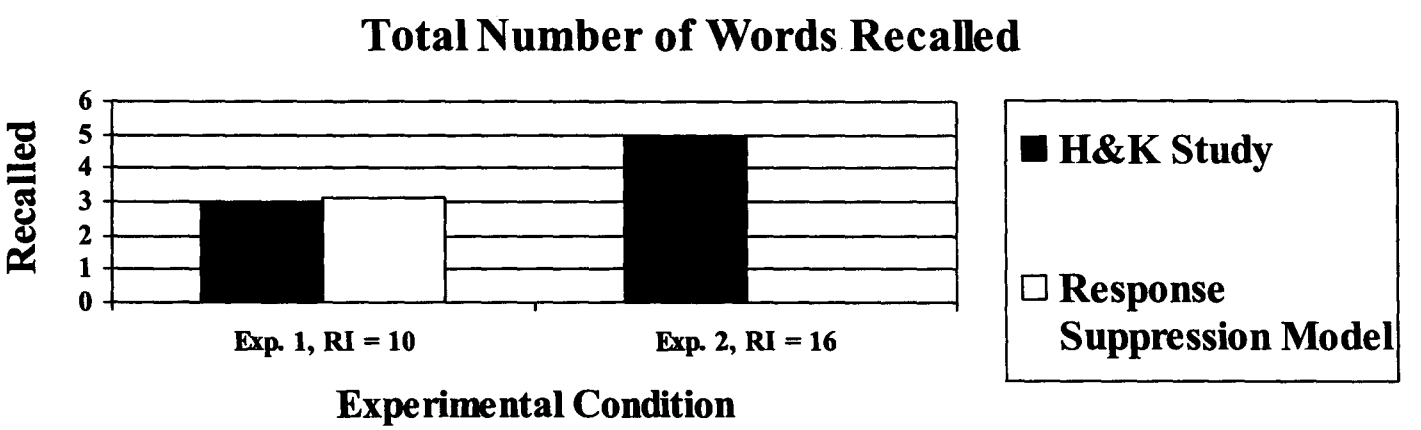

Total Number of Words Recalled
Recalled
6
5
4
3
2
1
0
Exp. 1, RI = 10
Exp. 2, RI = 16
Experimental Condition
H&K Study
Response Suppression Model

Figure 2.13

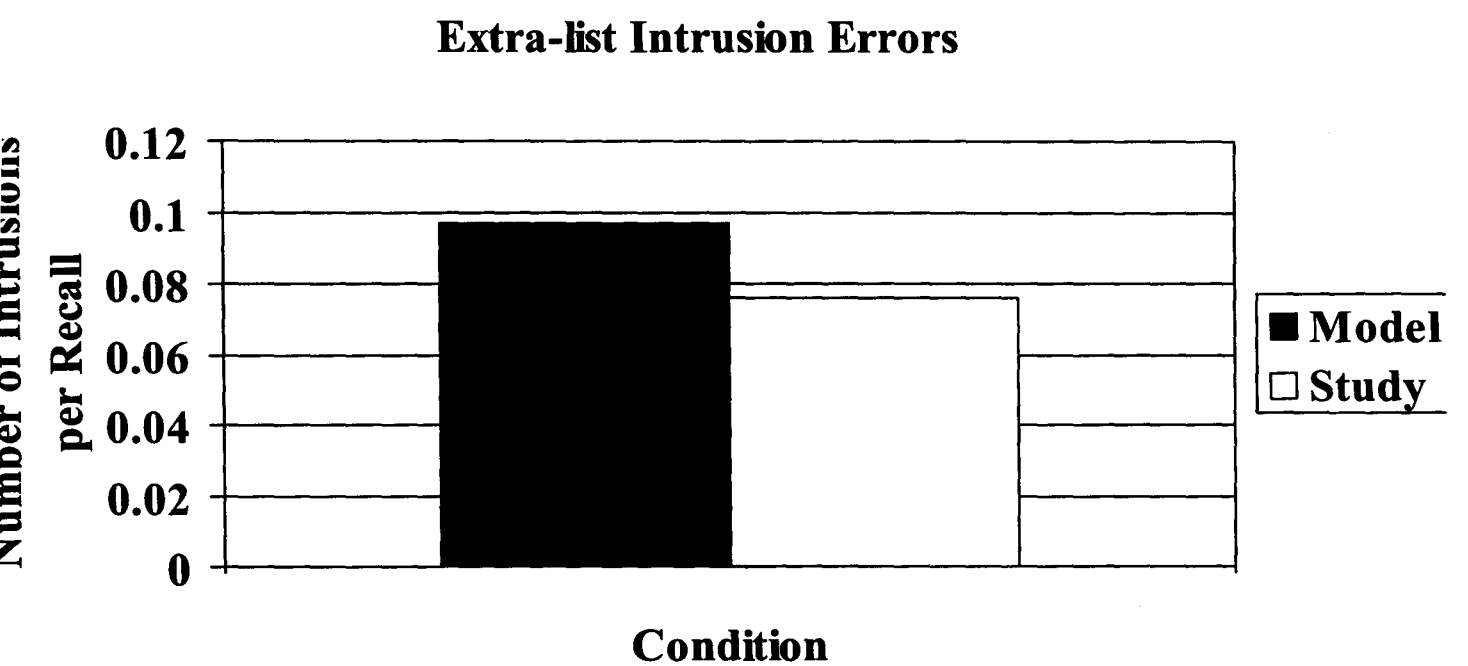

Extra-list Intrusion Errors
Number of Intrusions per Recall
0.12
0.1
0.08
0.06
0.04
0.02
0
Model
Study
Condition

Figure 3.1

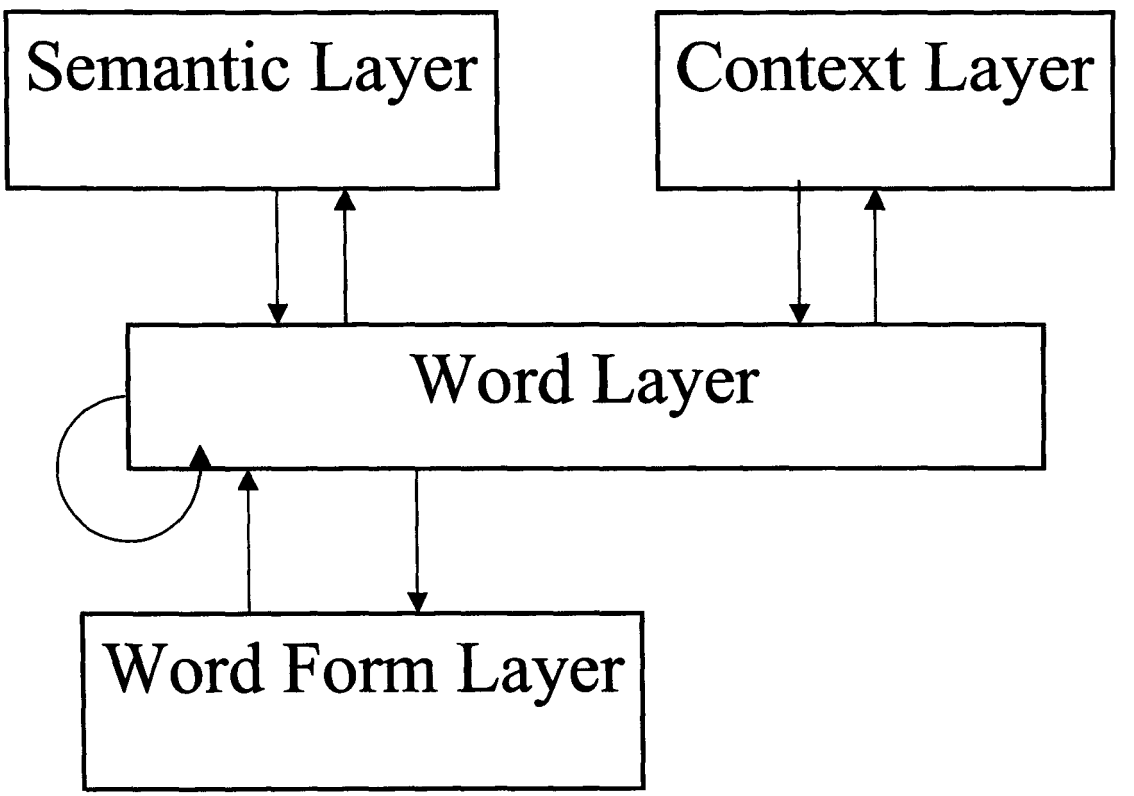


Figure 3.2

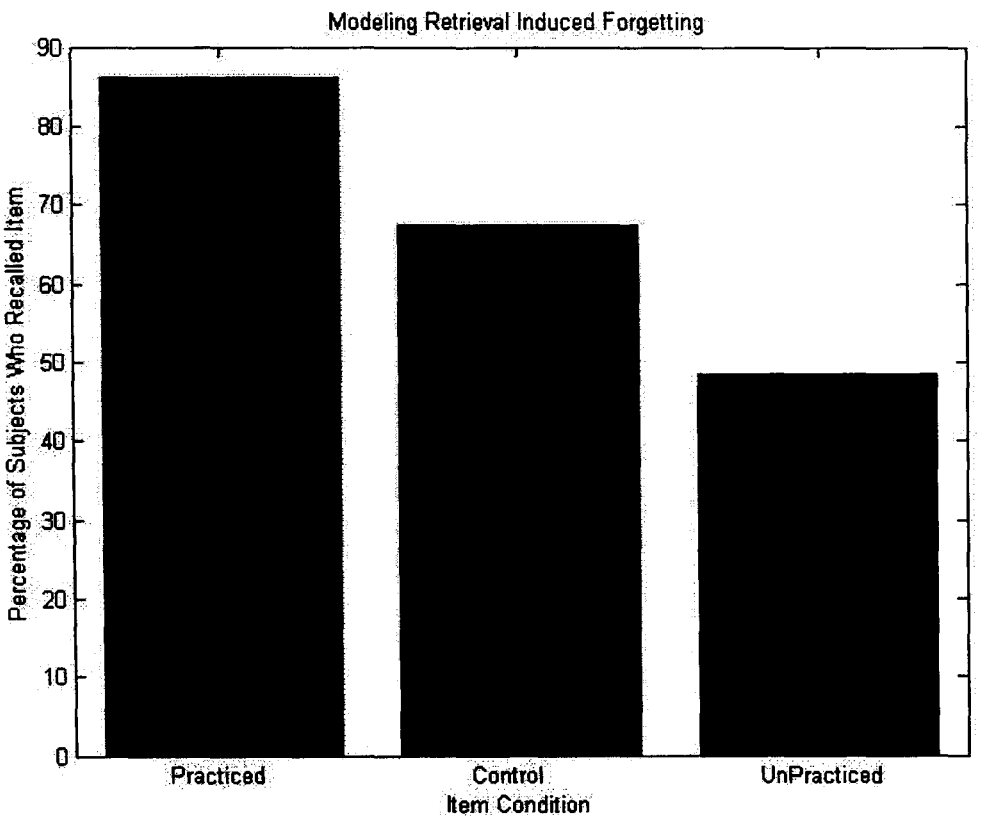

Modeling Retrieval Induced Forgetting
Percentage of Subjects Who Recalled Item
90
80
70
60
50
40
30
20
10
0
Practiced
Control
UnPracticed
Item Condition

Figure 3.3

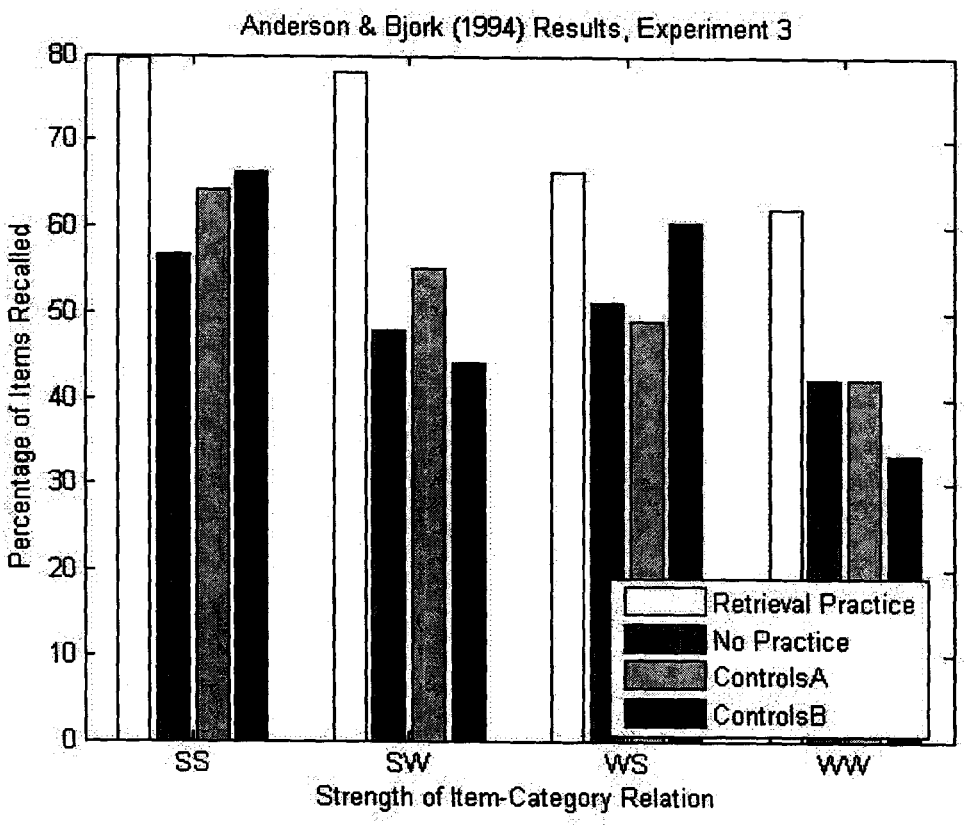

Anderson & Bjork (1994) Results, Experiment 3
Percentage of Items Recalled
80
70
60
50
40
30
20
10
0
SS
SW
WS
WW
Strength of Item-Category Relation
Retrieval Practice
No Practice
ControlsA
ControlsB

Figure 3.4

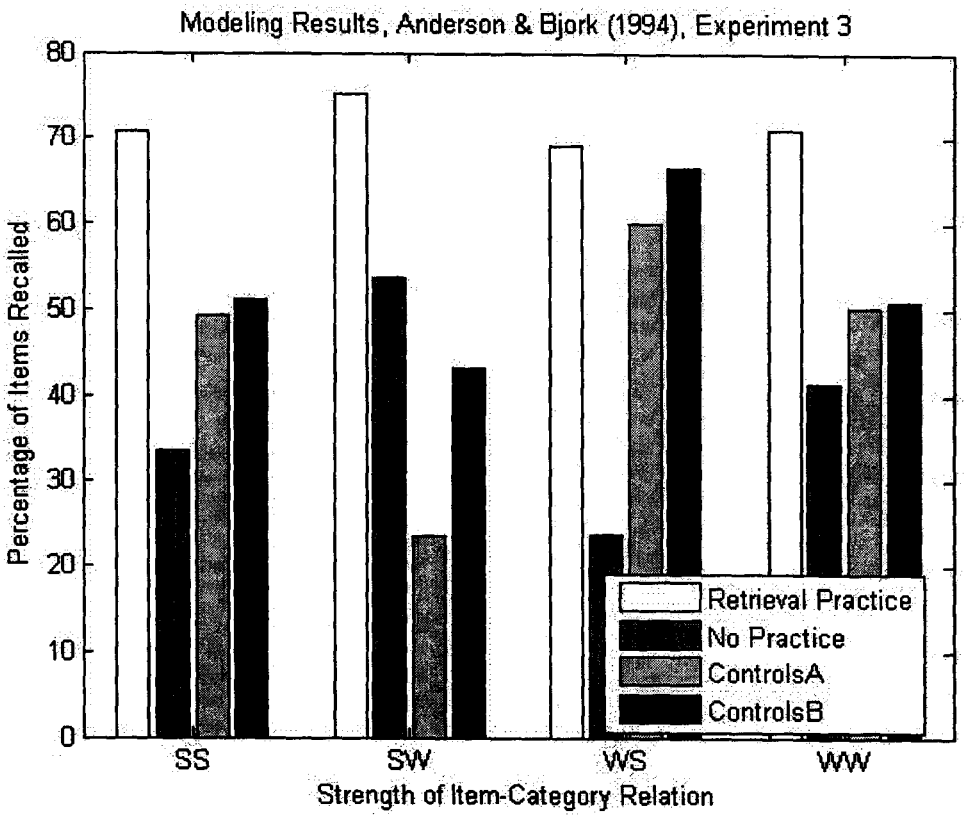

Modeling Results, Anderson & Bjork (1994), Experiment 3
Percentage of Items Recalled
80
70
60
50
40
30
20
10
0
SS
SW
WS
WW
Strength of Item-Category Relation
Retrieval Practice
No Practice
ControlsA
ControlsB

Figure 3.5

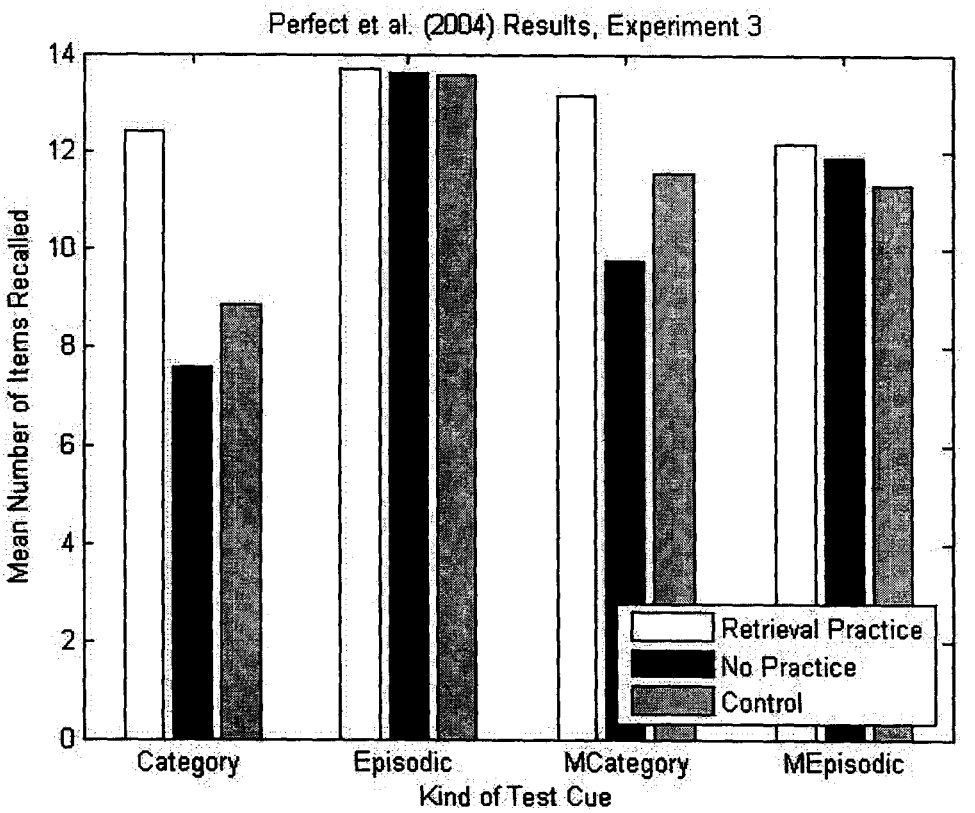

Perfect et al. (2004) Results, Experiment 3
Mean Number of Items Recalled
14
12
10
8
6
4
2
0
Category
Episodic
MCategory
MEpisodic
Kind of Test Cue
Retrieval Practice
No Practice
Control

Figure 3.6

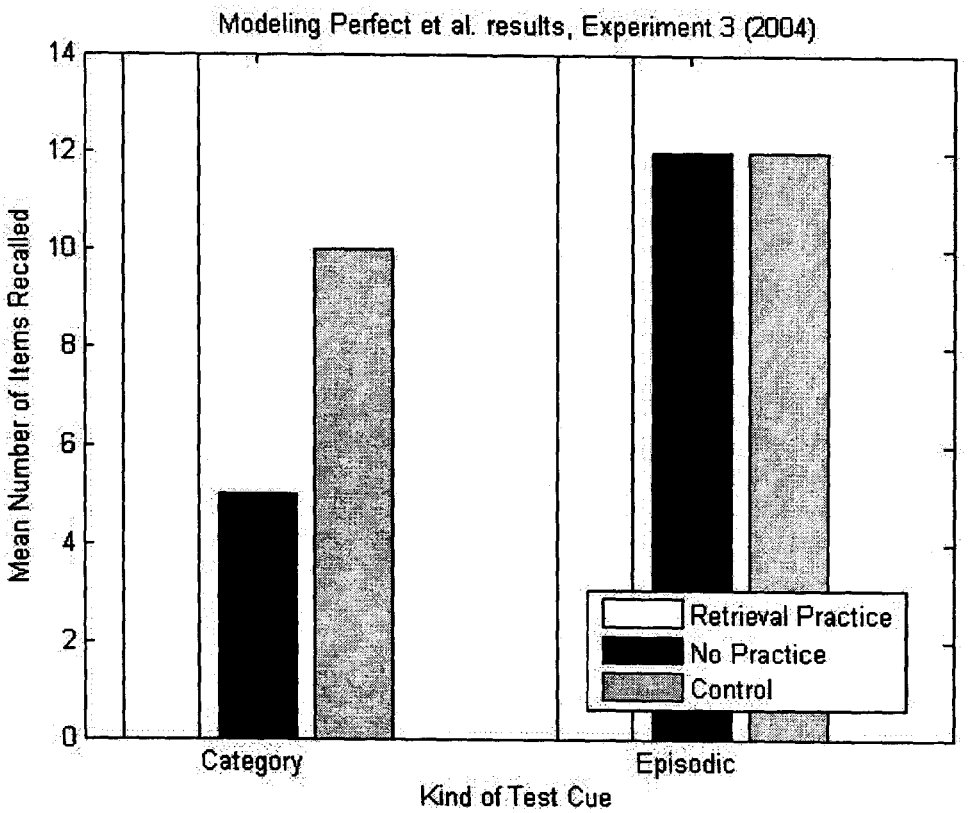

Modeling Perfect et al. results, Experiment 3 (2004)
Mean Number of Items Recalled
14
12
10
8
6
4
2
0
Category
Episodic
Kind of Test Cue
Retrieval Practice
No Practice
Control

Figure 3.7

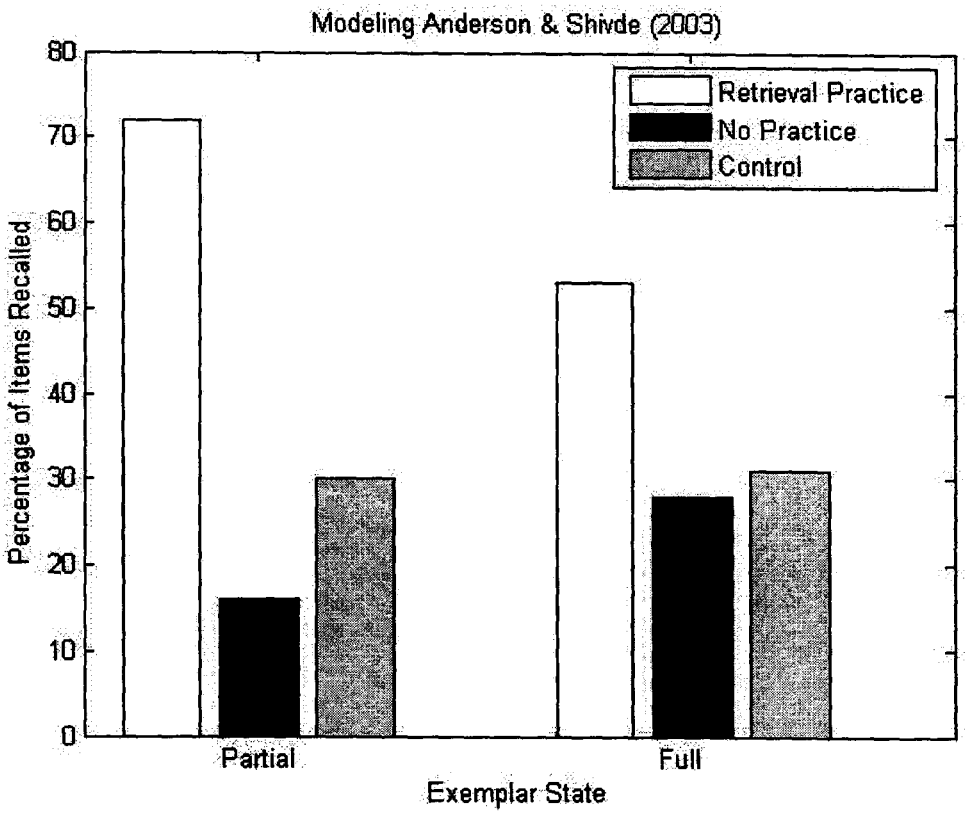

Modeling Anderson & Shivde (2003)
Retrieval Practice
No Practice
Control
Percentage of Items Recalled
80
70
60
50
40
30
20
10
0
Partial
Full
Exemplar State

Figure 4.1

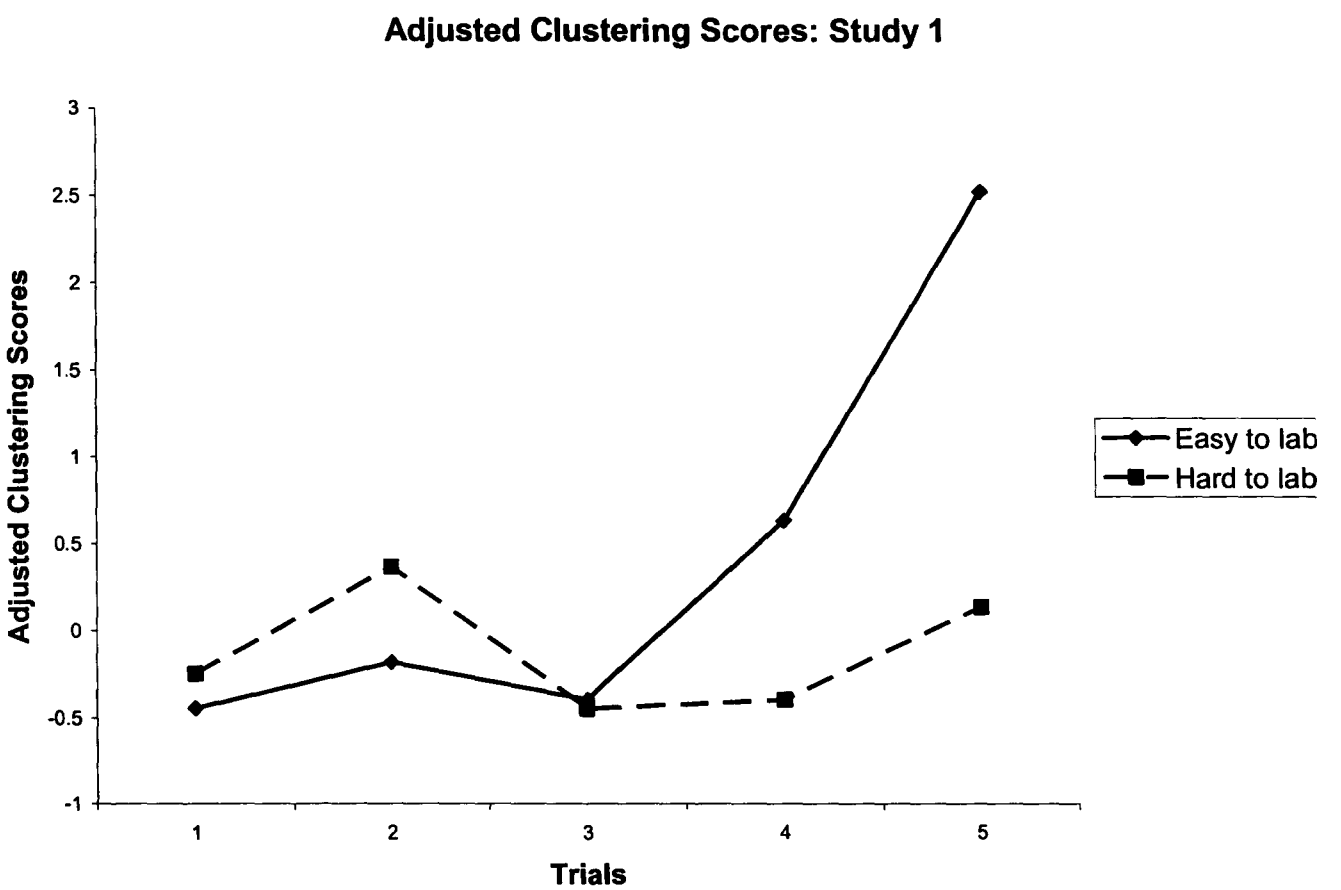

Adjusted Clustering Scores: Study 1
Adjusted Clustering Scores
3
2.5
2
1.5
1
0.5
0
-0.5
-1
1
2
3
4
5
Trials
Easy to lab
Hard to lab

Figure 4.2

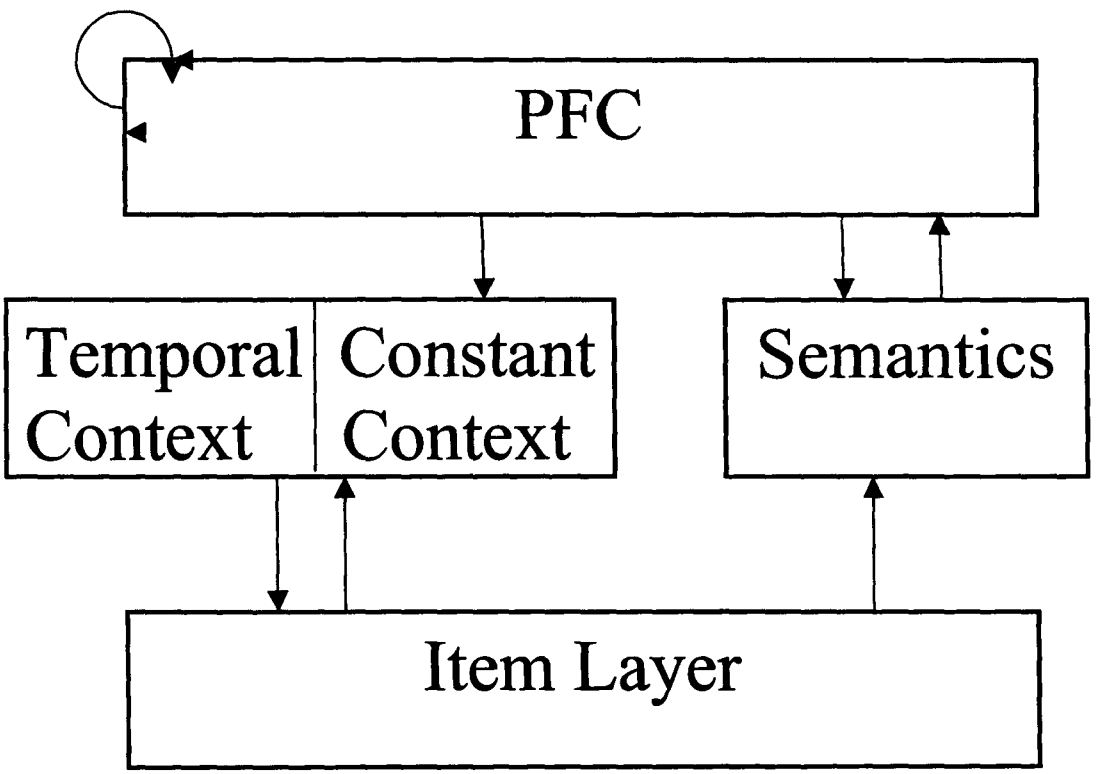

PFC
Temporal Context
Constant Context
Semantics
Item Layer

Figure 4.3

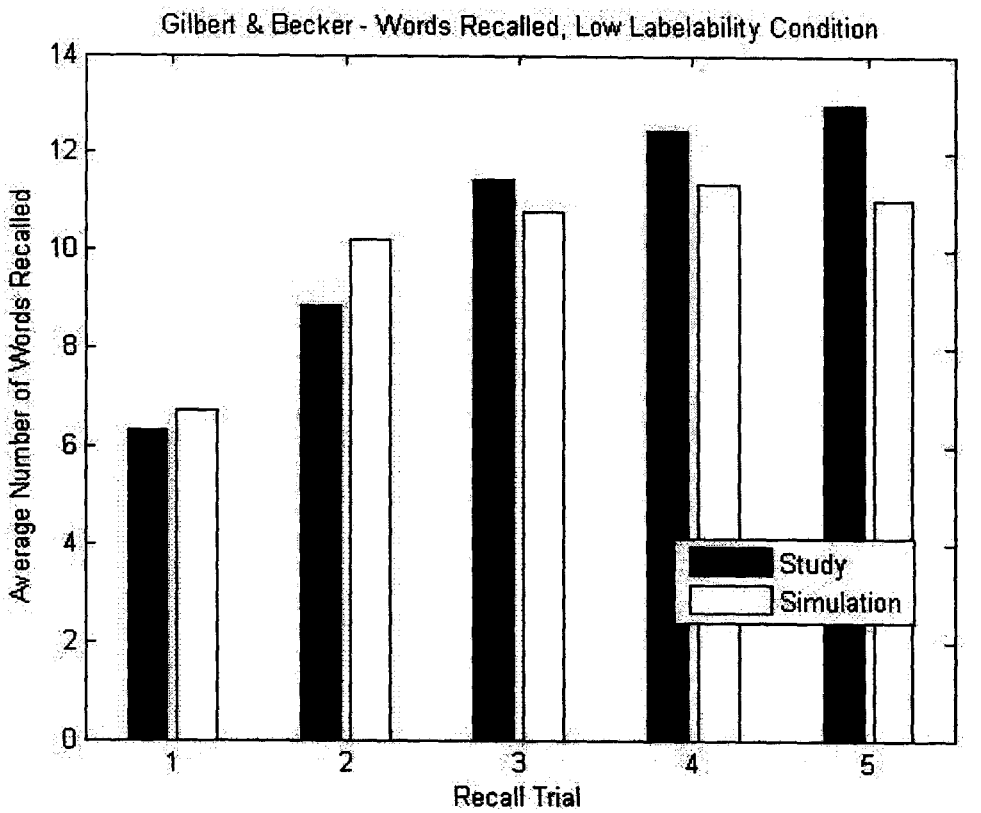

Gilbert & Becker - Words Recalled, Low Labelability Condition
Average Number of Words Recalled
Recall Trial
Study
Simulation
0
2
4
6
8
10
12
14
1
2
3
4
5

Figure 4.4

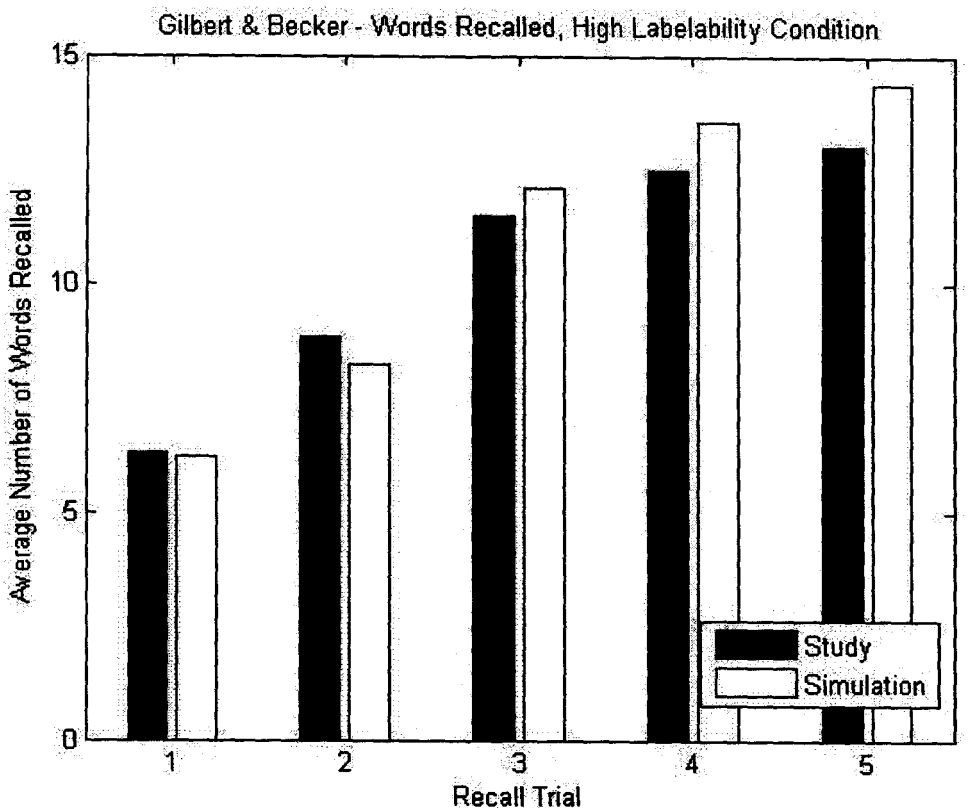

Gilbert & Becker - Words Recalled, High Labelability Condition
Average Number of Words Recalled
Recall Trial
Study
Simulation
0
5
10
15
1
2
3
4
5

Figure 4.5

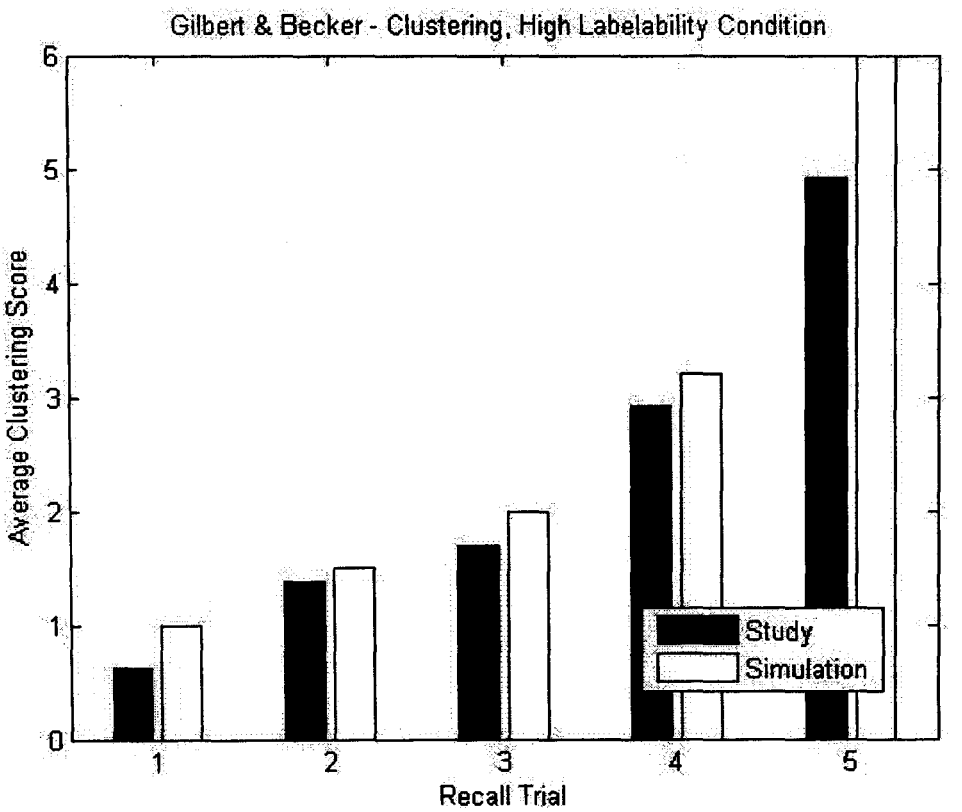

Gilbert & Becker - Clustering, High Labelability Condition
Average Clustering Score
Recall Trial
Study
Simulation
0
1
2
3
4
5
6
1
2
3
4
5

Figure 4.6

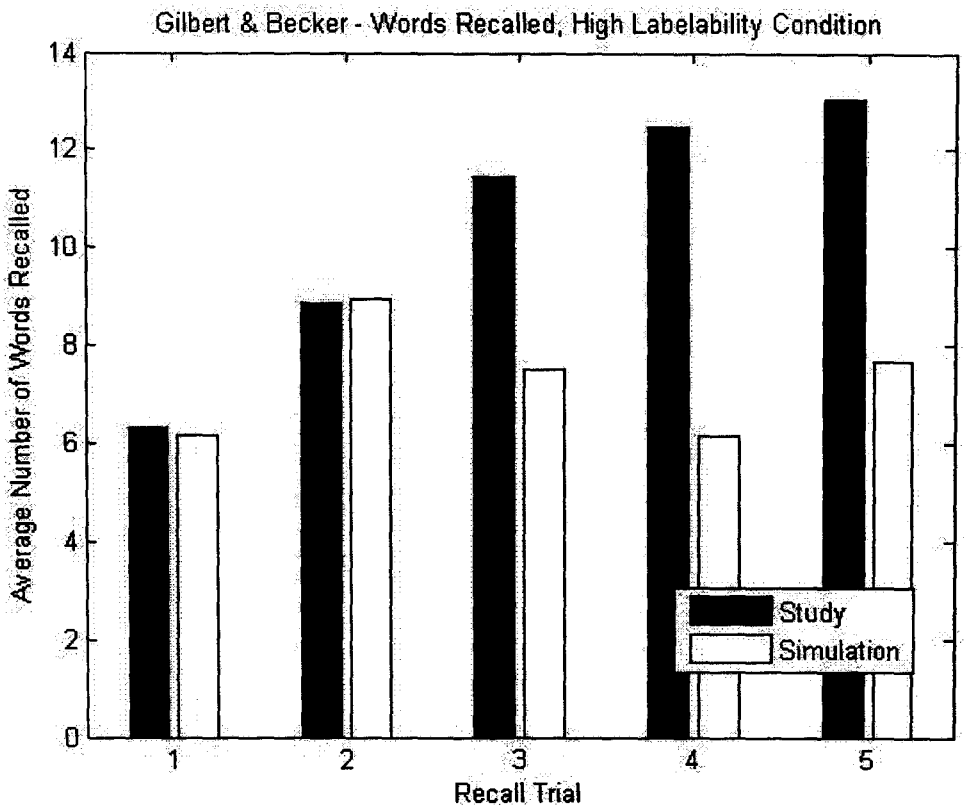

Gilbert & Becker - Words Recalled, High Labelability Condition
Average Number of Words Recalled
Recall Trial
Study
Simulation
0
2
4
6
8
10
12
14
1
2
3
4
5

Figure 4.7

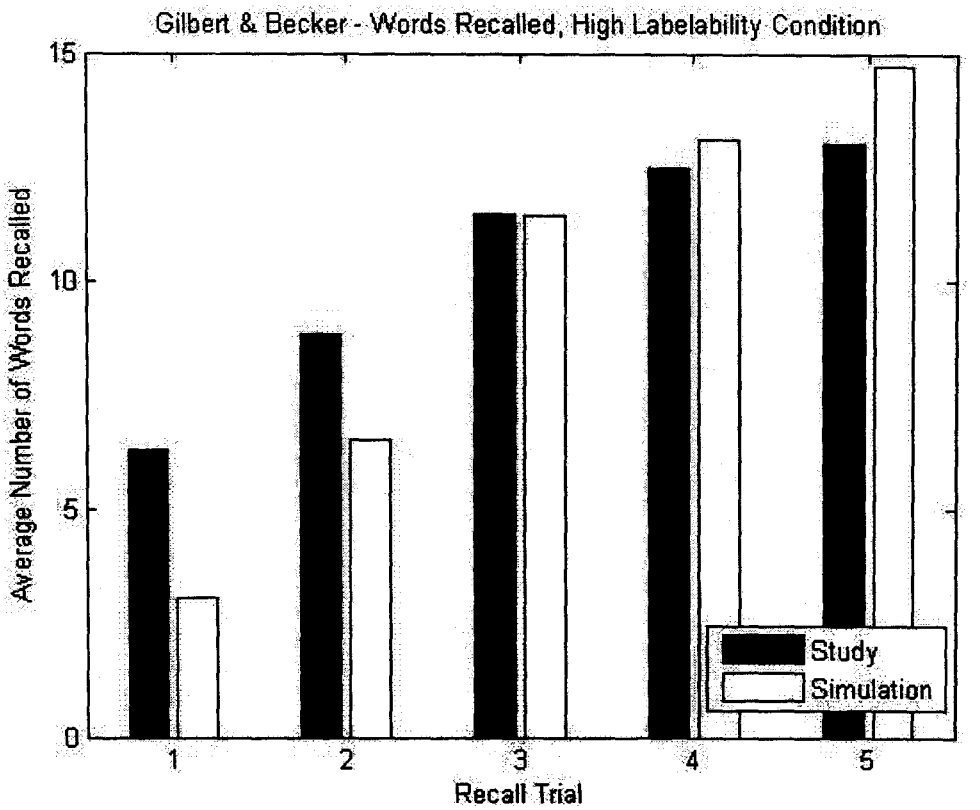

Gilbert & Becker - Words Recalled, High Labelability Condition
Average Number of Words Recalled
Recall Trial
0
5
10
15
1
2
3
4
5
Study
Simulation

Figure 4.8

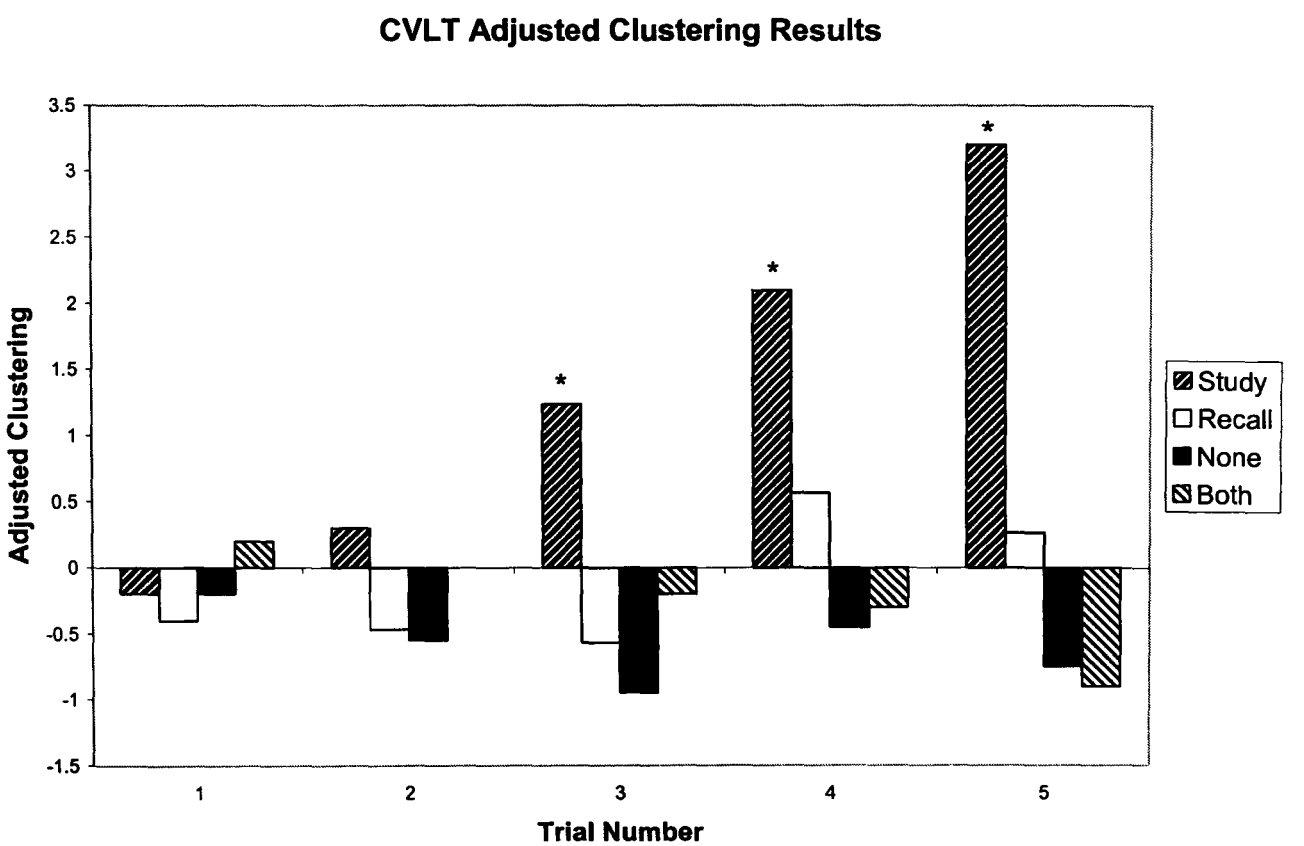

CVLT Adjusted Clustering Results
Adjusted Clustering
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
1
2
3
4
5
Trial Number
*
*
*
Study
Recall
None
Both

Figure 4.9

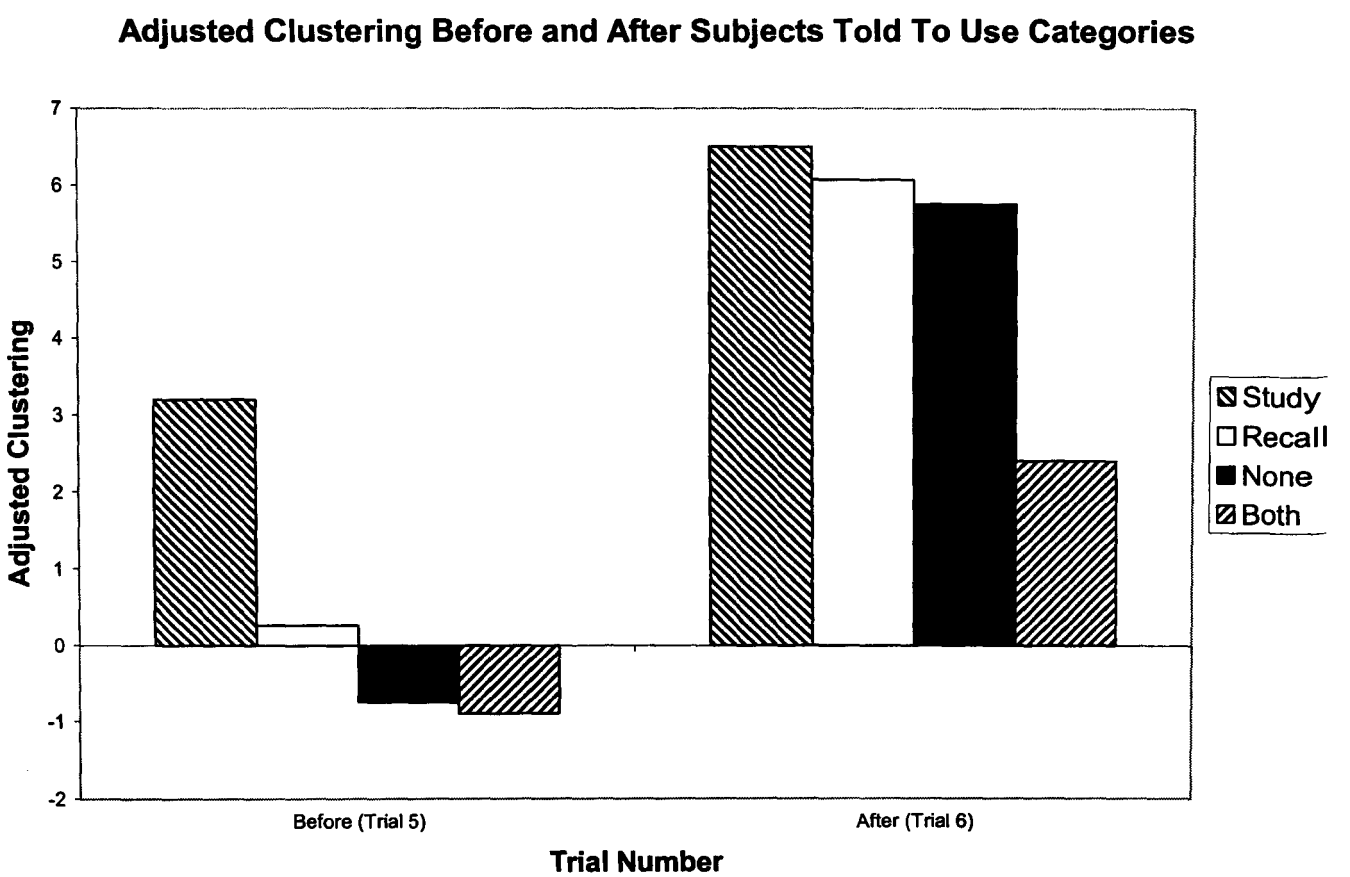
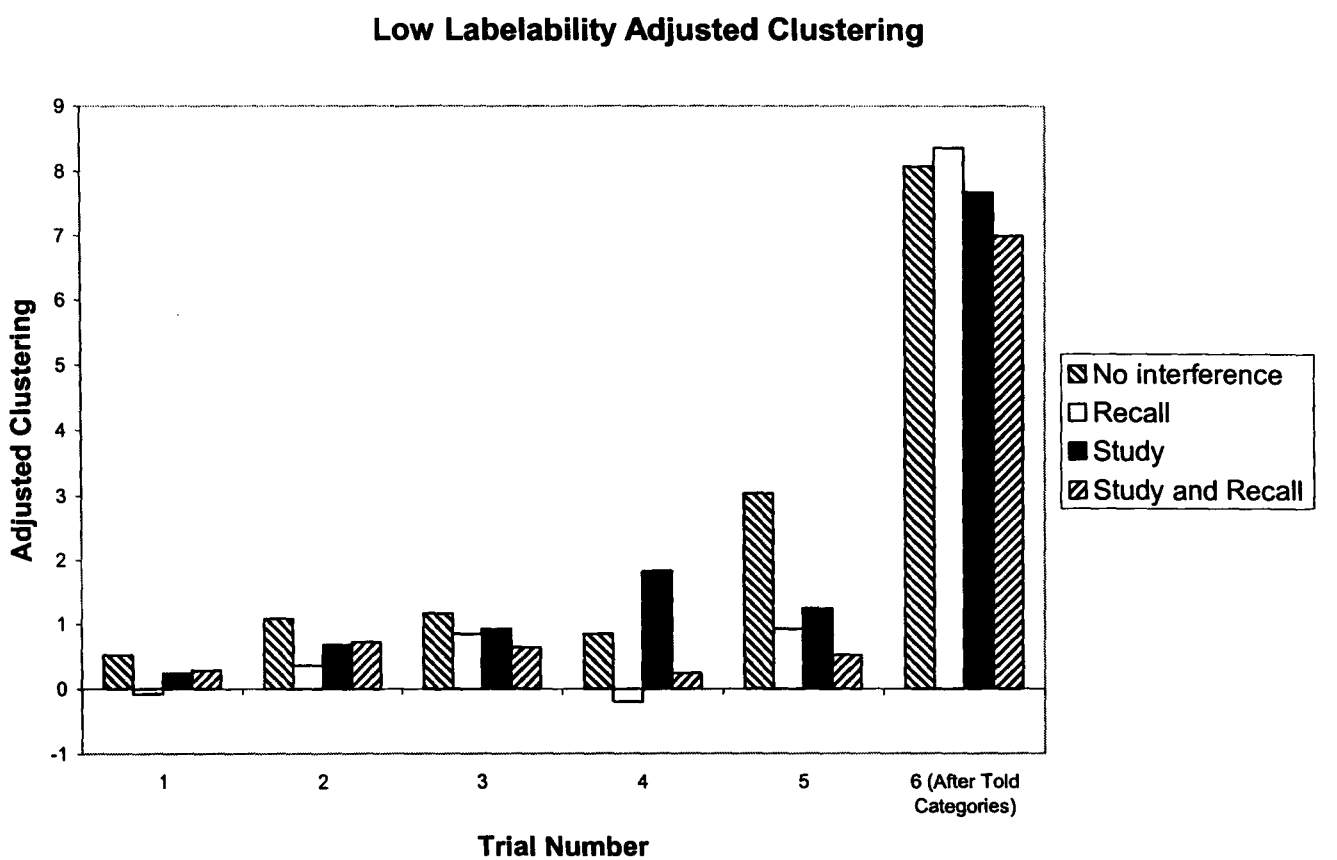

Adjusted Clustering Before and After Subjects Told To Use Categories
Adjusted Clustering
7
6
5
4
3
2
1
0
-1
-2
Before (Trial 5)
After (Trial 6)
Trial Number
Study
Recall
None
Both

Figure 4.10


Low Labelability Adjusted Clustering
Adjusted Clustering
9
8
7
6
5
4
3
2
1
0
-1
1
2
3
4
5
6 (After Told Categories)
Trial Number
No interference
Recall
Study
Study and Recall